Exploring flexibility and context dependency in the mycobacterial central carbon metabolism

DISSERTATION
zur Erlangung des akademischen Grades
DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

im Fach
Biophysik
eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl.-Ing. Katja Tummler

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. habil. Dr. Sabine Kunst
Dekan der Lebenswissenschaftlichen Fakultät
Prof. Dr. Bernhard Grimm

Gutachter/innen:
1. Prof. Dr. Dr. h.c. Edda Klipp
2. Prof. Dr. Uwe Sauer
3. Prof. Dr. Hermann-Georg Holzhütter

Tag der mündlichen Prüfung: 09.12.2016

Oh heavens, now,
is it the middle or right one to stop and the left one to go,
or middle to go, left to stop?

Maud Menten

Abstract

Tuberculosis remains one of the major global health threats responsible for over 1.5 million deaths each year. This 'success story' of the causative agent *Mycobacterium tuberculosis* is thereby closely linked to a flexible metabolism, allowing growth despite the restrictive conditions within the human host.
In this thesis, the flexibility of the mycobacterial central carbon metabolism is explored by modeling approaches integrating high-quality experimental data. The analyses zoom in from a network based view to the detailed functionalities of individual, virulence relevant pathways. The interconnection of the central carbon metabolism to the remaining metabolic network is charted as a prerequisite to characterize its thermodynamic landscape, debunking glycolysis as bottleneck in different nutritional conditions. Based on steady state metabolomics and proteomics data, regulatory sites for the metabolic transition between different carbon sources are predicted by a novel method. Finally, the flexible interplay between two seemingly redundant pathways for the catabolism of an *in vivo*-like carbon source is explained mechanistically by means of thermodynamic-kinetic modeling.
By employing novel modeling methods in combination with high-resolution experimental data, this work adds to the mechanistic understanding of the context dependent flexibility of mycobacterial metabolism, an important target for the development of novel drugs in the battle against tuberculosis.

Zusammenfassung

Tuberkulose ist auch heute noch eine der tödlichsten Infektionskrankheiten der Welt, der jährlich Millionen Menschen, vor allem in Entwicklungsländern, zum Opfer fallen. Ihr Erreger, *Mycobacterium tuberculosis*, unterwandert das menschliche Immunsystem durch eine Reihe von hochentwickelten Strategien, zu denen auch ein extrem anpassungsfähiger Stoffwechsel gehört. Die Dissertation untersucht diesen aus Sicht der Systembiologie und verbindet dabei Methoden der mathematischen Modellierung mit neuen, qualitativ hochwertigen experimentellen Daten. Ausgehend von einem Überblick über die metabolische Landschaft des zentralen Kohlenstoffmetabolismus und dessen systematischer Verknüpfung zu umgebenden Stoffwechsel- und Biosynthesewegen, erhöht sich Schritt für Schritt die Detailtiefe bis hin zur genauen, thermodynamisch konsistenten Analyse der kinetischen Eigenschaften spezieller infektionsrelevanter metabolischer Wege.

Basierend auf der Analyse der erhobenen experimentellen Daten werden zunächst allgemeine Größenordnungen für die Komponenten (Konzentrationen, Flüsse etc.) der nachfolgenden Modellierungsansätze etabliert. *M. tuberculosis* wird ausführlich "vermessen" und seine physiologischen Eigenschaften in Relation zu anderen Zellen gesetzt. Diese rein datenbasierte Auswertung gibt bereits wichtige Hinweise auf die Anpassung des zentralen Kohlenstoffmetabolismus unter verschiedenen Nahrungsquellen und Wachstumsphasen.

Diese solide Datenbasis wird verwendet um das zentrale metabolische Netzwerk *M. tuberculosis*' im Fließgleichgewicht zu charakterisieren. Zum einen wird eine neu entwickelte Methode vorgestellt, die solch isolierte metabolische Teilnetzwerke basierend auf einem bereits existierenden genomskaligen Modell sinnvoll in den Gesamtstoffwechsel einer Zelle einbettet ('Biomass Backtracking'). Neben einigen allgemeineren Anwendungsbeispielen wird die Methode verwendet um wachstumsbedingungsspezifische, biologisch korrekte Randflüsse zu errechnen, die für alle weiteren Modelle unabdingbar sind. Das kontextualisierte Netzwerk wird dann mit Hilfe eines Ansatzes der linearen Programmierung thermodynamisch charakterisiert. Dabei werden existierende Methoden, wie die thermodynamische Fluss-Bilanzanalyse oder die Optimierung der Maximum-Minimum driving force, kombiniert um thermodynamischer Optimalität auf Netzwerkebene Rechnung zu tragen. In diesem Licht stellt sich die Glykolyse als limitierender Stoffwechselweg in allen getesteten experimentellen Bedingungen dar.

Im nächsten Verfeinerungsschritt, wird das obige rein stöchiometrische System durch kinetische Gleichungen für alle beteiligten Reaktionen erweitert, um die Abhängigkeit der Flüsse von den experimentell quantifizierten Metabolitdaten zu ergründen. Das so entstandene Modell wird in einem neuen Optimierungsansatz verwendet um ein minimales Set regulatorischer Punkte zu identifizieren, das alle Unterschiede in den Metabolitdaten im Fließgleichgewicht wiedergeben kann. Dazu werden ebenfalls Proteindaten in das Modell integriert. Die Optimierung charakterisiert die obere Glykolyse sowie die zweite Hälfte des Zitronensäurezyklus als regulatorische Punkte für den Übergang zwischen glykolytischen und gluconeogenetischen Kohlenstoffquellen.

Im letzten Teil, wird das Fließgleichgewicht verlassen, um die Funktionalität zweier spezieller Abbauwege für Propionat, einer Kohlenstoffquelle die die Bedingungen im menschlichen

Wirt nachahmt, zu verstehen. Zusätzlich wird ein thermodynamischer Term in die kinetische Beschreibung der Reaktionen eingefügt. Basierend auf experimentellen Zeitverläufen der Konzentrationen der beteiligten Metabolite nach schnellen Wechseln der Kohlenstoffquelle, können den beiden scheinbar redundanten Abbauwegen spezifische Funktionen zugeordnet werden: Der thermodynamisch limitierte Methylmalonyl-Weg kann schnell auf Änderungen der intrazellulären Propionatkonzentration reagieren, auch um potentielle Intoxikation durch metabolische Zwischenprodukte zu verhindern. Der transkriptionell regulierte Methylzitrat-Weg, erreicht eine höhere Flusskapazität und metabolische Kontrolle, weshalb er als spezialisierter Abbauweg verwendet wird. Diese funktionelle Diversifizierung und mögliche Kompensation kann als ein Konzept der krankheitsrelevanten metabolischen Flexibilität *M. tuberculosis*' gesehen werden.
Durch die Weiterentwicklung von Methoden zur metabolischen Modellierung sowie die gezielte Integration hochauflösender, quantitativer Metabolit- und Proteomdaten trägt die Arbeit so ihren Teil für ein besseres Verständnis der Grundlagen des zentralen Kohlenstoffwechsels in Mykobakterien bei.

Contents

A Appendix

Introduction

1. Introduction

Are you interested in tuberculosis? The following Chapter will introduce basic facts on tuberculosis and explain why studying the pathogen *Mycobacterium tuberculosis*, and especially its metabolism, is at the core of efforts to tackle the global tuberculosis pandemic. It will also describe why integrative systems biological approaches – such as mathematical modeling and large scale experimental measurements – are required to advance understanding in this field.

1.1 Tuberculosis

1.1.1 Epidemiology and pathogenesis

A forgotten enemy

While tuberculosis incidences in Europe are almost negligible and have hence vanished from public attention, the disease remains a major global health threat: Around one third of the world's population is infected with *M. tuberculosis*. Even though infected patients have an only 10 % lifetime risk of developing open tuberculosis, the active form of the disease, the death toll accounts to over 1.5 million persons per year, or one each 21 seconds. Only second to HIV, tuberculosis is a leading cause of premature death due to infection worldwide. Incidences are high in developing countries, with 22 high-burden countries registering over 80 % of the global tuberculosis cases (Figure 1.1).

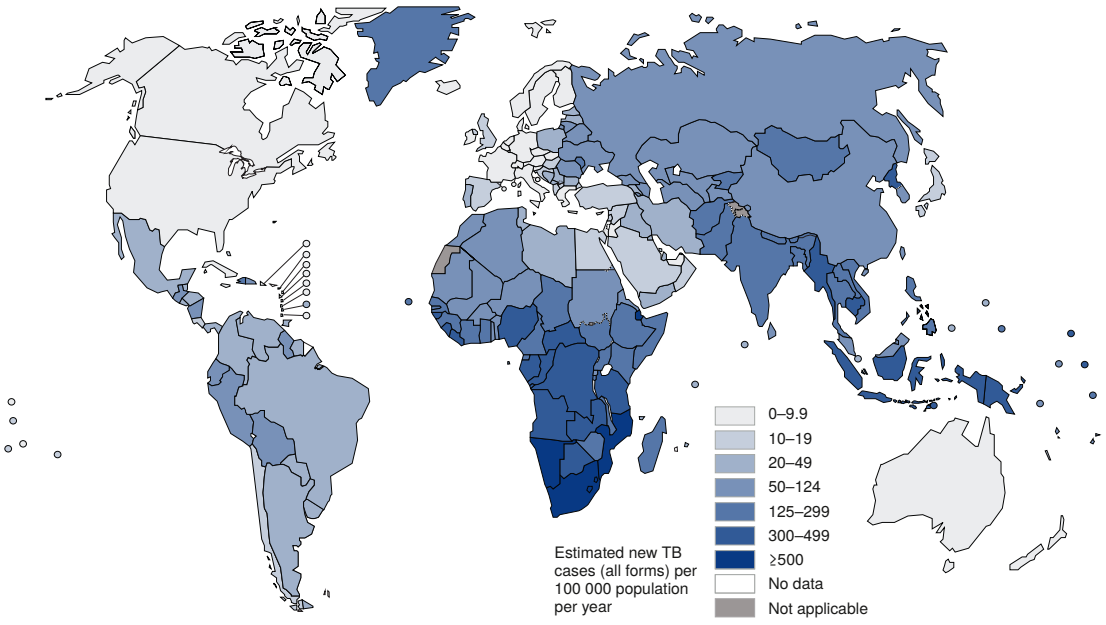


Figure 1.1: Tuberculosis incidence per year and country, darker blue indicates a higher number of tuberculosis patients. The figure was taken and modified from the WHO global tuberculosis report [209].

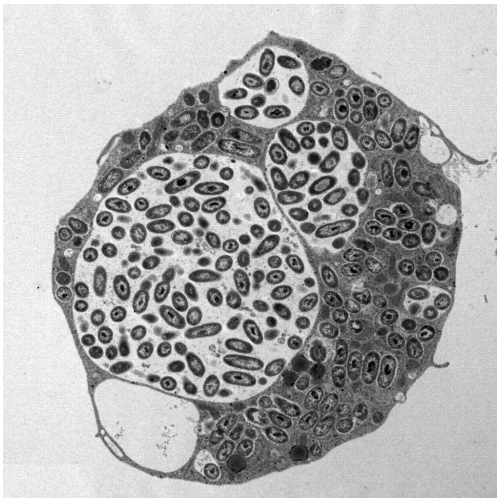


Figure 1.2: Alveolar macrophage with phagocytized *M. tuberculosis*, isolated from a patient (electron micrograph). The figure is taken from Russell *et al.* [167].

Access to medication and diagnostics is often limited in those countries, explaining the fatalities despite the availability of effective antibiotic medication. Nonetheless, there are also success stories to be told: Global efforts focused on improved diagnosis and treatment have decreased tuberculosis mortality by 47 % and incidences by more than 18 % in the last 25 years. Newly developed drugs with shorter treatment regimens can further advance this progress. [209]

Pathogenesis

After having co-evolved with its human host, probably since the beginnings of modern mankind [61], *M. tuberculosis* was identified as the causative agent of tuberculosis by Robert Koch in 1882 [95]. Due to this co-evolution, *M. tuberculosis* is a highly specialized pathogen, perfectly adapted to its living environment. A solid body of knowledge on many aspects of tuberculosis' pathogenesis, *M. tuberculosis* itself and its interplay with the host has accumulated. Here, a concise summary of the aspects relevant to this work will be given.

M. tuberculosis enters (and leaves) the human host via the respiratory tract in small aerosol droplets. The infectious dose is thereby as low as 10-100 bacteria, rendering tuberculosis an extremely contagious disease. The locus of the primary infection is usually in the lower regions of the lung, whereas in later stages of the disease, *M. tuberculosis* can invade virtually all other tissues in the body.

The bacteria hijack an integral compound of the human immune system, the ***macrophage***. These professional phagocytes engulf the bacteria within phagosomes (Figure 1.2). By acidification of the phagosomal lumen and the secretion of reactive oxygen and nitrogen intermediates, the macrophages in the most cases clear engulfed bacteria. However, during the co-evolution with its human host, *M. tuberculosis* has learned to hinder this phagosomal digestion process at several levels to conquer the phagosome as its habitat. Secreted proteins inhibit the acidification of the phagosome as well as its fusion to the lysosome – both steps that are required for bacterial killing. The arrest in phagosome maturation also reduces the contact to the antigen processing machinery required for an efficient adaptive immune response. At the same time, the *M. tuberculosis*-bearing phagosomes retain fusion with recycling endosomes containing a limited pool of nutrients, which can be exploited for bacterial growth. Within this harsh environment, the bacteria grow slowly, with division times around 24 h. The scarcity of nutrients entails the need for an extremely flexible metabolism, which can utilize small amounts of distinct nutrients. Also the physiology of the bacteria contributes to coping with the hostile

conditions in the phagosome: *M. tuberculosis* is shielded by a thick cell envelope, consisting of specialized lipids, the eponymous mycolic acids and others, which act as major virulence factors and nutrient storage at the same time. [87, 88, 167, 174]
M. tuberculosis-infected macrophages, in their struggle to kill the internalized bacteria, are imprisoned by the host immune system in so called ***granulomas*** – localized aggregates of infected macrophages tightly surrounded by lymphocytes. The granuloma mainly acts as a local constriction of the bacteria, as a restriction to nutrient and oxygen availability and as a focal point of the immune response. Nevertheless, the host immune system often fails to clear the granuloma from *M. tuberculosis*, such that the bacteria survive within it for long time periods, even for the entire lifetime of the infected host. *M. tuberculosis*, on the other hand, has learned to exploit its prison, infecting new immune cells that are delivered directly to the contained bacteria. Additionally, the bacteria modulate the granuloma interior to reduce adaptive immune cell proliferation. Complex bacterial lipids are at the core of this process, triggering the release of inappropriate cytokines from host cells. [167]
In this see-saw battle between host and pathogen, *M. tuberculosis* exits its cell cycle if living conditions become too harsh. The bacteria enter a state of ***dormancy***, characterized by drastically reduced metabolic activity, absence of growth and tolerance to antibiotics. Dormant bacteria are associated with phases of latency of the disease, sustaining the danger of re-activation over decades, especially in situations of reduced immunocompetence. [49, 63, 89, 187]

1.1.2 Frontiers in tuberculosis treatment

Tuberculosis can be treated with antibiotics, but due to the slow growth rates, the inaccessible micro-environments in the granuloma and the presence of dormant bacteria with low uptake of drugs, treatment regimens are long (6 months) and require four antibiotics at the same time.
If the treatment is interrupted or not completed, subpopulations of ***resistant*** *M. tuberculosis* can evolve. In countries with poor health-care systems and infrastructure, resistant strains are frequent, especially in previously treated patients. Treatment against resistant strains requires taking more expensive drugs with larger side effects for longer periods (up to 20 months) and has a lower success rate. This clearly highlights the need for the development of new drugs with the capability of bypassing the evolved resistance mechanisms. [187, 209]
Especially in African countries, ***co-infection with the human immunodeficiency virus*** (HIV) impedes tuberculosis control. Immunocompromised patients infected with HIV and *M. tuberculosis* have a drastically higher risk to develop open tuberculosis and treatment outcomes are worse. Co-infection speeds up the progress of either disease and impedes tuberculosis diagnosis. Additionally, patients are more susceptible to resistant strains, as mutated strains are usually less fit and can often be cleared by immunocompetent hosts. Again, more effective antibiotics, but also a better understanding of the mechanistic processes in *M. tuberculosis*' biology can help to address these problems. [2, 209]
Although not directly linked to this work, ***vaccination*** as a major frontier in global tuberculosis control is mentioned here in brief. External activation of macrophages by the adaptive immune response overwrites the phagosome maturation block in infected macrophages and is an efficient way to clear bacteria – vaccination can hence be a powerful tool in the battle against tuberculosis. However, the only available vaccine, an attenuated form of *M. bovis* BCG which causes bovine tuberculosis, shows poor efficiency, especially in adults. A number of new vaccines is under development, representing a major route on the way to the eradication of the disease. [87, 187, 209]

1.2 Metabolism

When focusing on the pathogen side of tuberculosis, the mycobacterial metabolism is a promising research target, as it is central to a multitude of virulence mechanisms: The synthesis of complex lipids plays an important role in immunomodulation and physical protection of the bacteria, but inflicts a high burden on the metabolism. The nutrient limitation in the phagosome requires an extremely flexible metabolism that can utilize non-canonical carbon sources in small quantities. Transition to the dormant state requires an almost complete shutdown of metabolism, but at the same time residual metabolic activity has to be guaranteed for resuscitation, potentially after decades of dormancy. As all those metabolic processes are very distinct from the metabolism in the human host, they are promising targets for novel antibiotics.
In the following Section, fundamentals on cellular metabolism will be introduced and subsequently linked specifically to the metabolism in mycobacteria.

1.2.1 Complexity, modularity and regulation

Fundamentals

Metabolism is the set of chemical reactions within a cell that facilitates the production of energy and biomass required for cell growth from external nutrients. The nutrients are thereby broken down by catabolic reactions into a conserved set of precursor metabolites. These precursors are then used in anabolic reactions to build the integral parts of the cell's biomass: Proteins, nucleic acids, lipids and complex sugars. This so-called ***bow-tie structure*** allows for the same set of chemicals to be used to build diverse complex macromolecules, sharing many steps on the way to their production. Such production ways are referred to as ***metabolic pathways*** and represent a set of subsequent chemical transformations. The entirety of pathways in a cell constitutes its metabolic network, where reactions are inter-wired and inter-dependent on many others.
Besides the production of precursors, the breakdown of extracellular nutrients in catabolic pathways additionally releases energy, which can be stored by the cell in form of high energy compounds such as adenosine-triphosphate (ATP) or as redox equivalents. This energy is then used in anabolic pathways to build complex molecules from the precursors.
Metabolic reactions are usually facilitated by ***enzymes*** – cellular proteins that act as catalysts. Enzymes speed up the chemical reaction by orders of magnitude and are also effective levers to control the metabolic activity in the cell. Enzymes are subject to evolutionary pressure: The better catalysts they are and the less material needs to be invested in them, the faster the cells can grow, resulting in an evolutionary advantage over other organisms in the same habitat.
Metabolism is tightly linked to other cellular processes via regulatory paths. Often metabolites are sensors for the overall state of the cell and accordingly adapt other processes like the cell division cycle, the replication of the genetic information, mating or the transition to dormancy.

What shapes metabolic flux?

As environmental conditions are usually changing over time, but metabolism always has to guarantee for the production of functional biomass, metabolic adaptation is required. Changing availability of nutrients but also environmental stresses result in differential utilization of metabolic pathways, the ***flux*** v of metabolites through specific reactions is altered. To achieve this re-routing of mass flux, several factors can be tuned by the cells.

Conceptually, v can be described as a function of five major contributions:

$$v = f\Big(\underbrace{\text{enzyme amount}}_{\text{(i)}}, \underbrace{\text{catalytic activity}}_{\text{(ii)}}, \underbrace{\text{enzyme saturation}}_{\text{(iii)}}, \underbrace{\text{thermodynamic gradient}}_{\text{(iv)}}, \underbrace{\text{regulation}}_{\text{(v)}}, \cdots\Big). \tag{1.1}$$

First and foremost, the flux depends on the availability of catalyzing enzyme (i), but also the catalytic properties of the enzyme influence the flux. Two parameters are thereby relevant: The binding affinities of the metabolites to the enzyme and the catalytic activity of the enzyme once a substrate is bound. In simpler terms, it is important how readily enzyme and substrate bind to each other and how much the catalysis speeds up the underlying chemical reaction. The latter term directly translates to increasing fluxes (ii), whereas the former changes with the concentration of available substrate metabolites, defining the saturation state of the enzyme (iii). More specifically, if a high concentration of a high-affinity substrate is present, all available enzyme molecules will constantly be occupied by substrate molecules, resulting in a saturated enzyme population without idle enzymes.
Chemical reactions also have an energetic aspect: According to the rules of thermodynamics, they can only occur spontaneously if they dissipate energy. The more energy a reaction dissipates, the more likely it will take place. This aspect affects metabolic fluxes via the thermodynamic gradient (iv) which depends on the chemical nature of the reaction as well as on substrate and product concentrations.
Finally, cells have developed regulatory mechanisms (v) to further alter the flux through a reaction. They can either affect the enzyme or interfere with the binding of substrates. In the former case, specific substances can additionally bind to the enzyme or compete with the substrate for binding sites, to activate or inhibit the reaction. In the latter case, the enzyme itself is modified, for example post-translationally by phosphorylations, which enhances or decreases the catalytic activity. Often, these regulations are the outcome of signaling pathways that transmit environmental information into the cell.
The above factors can be seen as constraints but also as levers for the cell to adapt metabolic fluxes according to its environment. All five are addressed in this thesis to characterize metabolism in mycobacteria (see Section 1.4). Although covering the major flux shaping factors in bacteria, this is not an exhaustive analysis, as further conceivable aspects can contribute to flux control, for example enzyme localization in the cell or co-localization with other enzymes. [46, 93, 136]

Central carbon metabolism

The central carbon metabolism is the backbone of cellular catabolism which converts sugars into biomass precursors. It includes the well studied pathway of glycolysis, the pentose phosphate pathway and the Krebs cycle, along with a set of connecting reactions. It is surprisingly well conserved between species [139], potentially because it represents an evolutionarily beneficial minimal set of chemical reactions for its task [135]. Central carbon metabolic enzymes often take up a large fraction of the cellular protein content, especially in bacterial cells [105], and have higher catalytic capacities compared to remote pathways [11], corresponding to the large fraction of the total metabolic flux that is processed by them. Alterations in the environmental conditions directly affect this part of metabolism, as catabolic routes of almost all nutrients converge here. Especially beneficial for quantitative as well as theoretical studies, the central carbon metabolism is well accessible to experimental measurements, allowing for time resolved determination of metabolite concentrations and reliable quantification of involved enzymes. For these reasons, the central carbon metabolism was chosen as a target for this work.

1.2.2 Mycobacterial metabolism

As described above, mycobacterial metabolism is a promising frontier in the battle against tuberculosis. However, many open questions remain, including the precise wiring of the central metabolic network, preferentially used enzymes and even the major carbon source *in vivo* [158]. Finding the major carbon source(s) in the pathogenic context, and alongside the active and essential pathways, is central to identifying promising drug targets. A body of evidence has accumulated for the primary utilization of host derived lipids, especially cholesterol [99, 127, 142]. This is consistent with the genes of β-oxidation, the major catabolic route for fatty acids, being heavily duplicated in the *M. tuberculosis* genome [37] and up-regulated during infection [171]. In addition, other studies measured carbon dioxide fixation [21] and predicted the metabolization of amino acids and other compounds containing two to three carbon atoms [22] by intracellular *M. tuberculosis*. The utilization of the latter carbon sources, as for example acetate or pyruvate, has been shown to contribute to the ability to grow in low pH media [8], mimicking the intraphagosomal environment.
How these potential nutrients could be taken up by the bacteria is poorly understood [132], but certainly challenged by the thick cell envelope. Downstream of nutrient import, most genes of the canonical central carbon metabolism are present in *M. tuberculosis*, but their exact wiring is debated [127, 194, 203]. Nevertheless, mutagenesis studies identified many of the genes as essential for mycobacterial growth [170] and infection [171].
Dormant bacteria are almost impossible to kill by antibiotics and therefore form a natural reservoir for tuberculosis. However, even in the dormant state, the bacteria show residual metabolic activity [24, 64]. Characterizing active metabolic functions in this context can reveal drug targets, whose disruption specifically eliminates the non-replicating bacterial population.

1.3 Systems, models and data

The previous Sections outlined the importance of studying *M. tuberculosis* and its metabolism, but also the complexity of the system became evident. Especially the central metabolic pathways are highly interconnected. They depend on environmental conditions and are subject to cellular regulation. In addition, the mere size of the system renders it beyond intuitive understanding. It is, for example, difficult to understand how the lack of canonical carbon sources in the phagosome influences cellular growth and energetics. Are fatty acids enough to compensate for them? What happens to the redox state of the cell, will there be enough energy for gluconeogenesis?
In the following Section, the integration of mathematical modeling with experimental measurements is described as means for studying such systems on different levels of granularity, while combining a multitude of available information.

1.3.1 More than the sum of it's parts

Complex systems, one of which is cellular metabolism, have the striking property that their behavior cannot be understood by examining their constituents individually. The interactions between the constituents give rise to ***systemic features***, such as robustness, memory, compensation, amplification and others. Such effects cannot be achieved on the level of individual agents, but are integral features of live. To add even more complexity, cellular metabolism is tightly connected to other cellular processes, such as gene expression, protein production or cellular signaling, which add additional complexity.

Computer aided mathematical modeling is a way to analyze systemic behaviors, allowing for the processing of large numbers of system constituents, linked by complex mathematical rules. Those rules are abstractions of the underlying biological mechanisms. In ***systems biological studies***, such as the one presented here, mathematical models are equipped with and tailored to available biological data in order to learn about mechanistic processes within cells. [93, 165]

1.3.2 Big data! Big models?

Brave new omics

Advances in experimental biology and measurement techniques in the last decades have made it possible to measure a vast amount of intracellular compounds. The entire genetic information of a cell, the genome, can be accessed and modified using genomics, proteomics identify and quantify all expressed proteins, the pool of transcribed genes is captured by transcriptomics and intracellular metabolites can be measured by metabolomics. Besides their grandiosity, these so-called ***omics*** approaches share their striving for completeness: The entirety of the cell is to be examined. Omics efforts generated a plethora of large datasets over the past years, that have revealed many facts about the functional composition of cells in different environmental conditions. However, these datasets are often too complex to intuitively draw mechanistic conclusions from them, such that many omics studies remain at a mostly descriptive level.

Models as useful abstractions

Mathematical ***models*** come into play to understand the mechanisms behind the measured data. Models are formalizations of knowledge obtained from different sources, including measured data as well as abstract concepts. Metabolic models, for example, rely on data about the concentrations of metabolites and enzymes, but also include concepts of mass balance, thermodynamics and the mechanisms of enzyme action. The formalization is based on (often nonlinear) mathematical equations, which approximate the dependencies of the participating agents. For example, each of the terms in Equation 1.1, can be described as a mathematical function expressing the dependency of the flux on metabolite and enzyme concentrations, kinetic parameters and others.

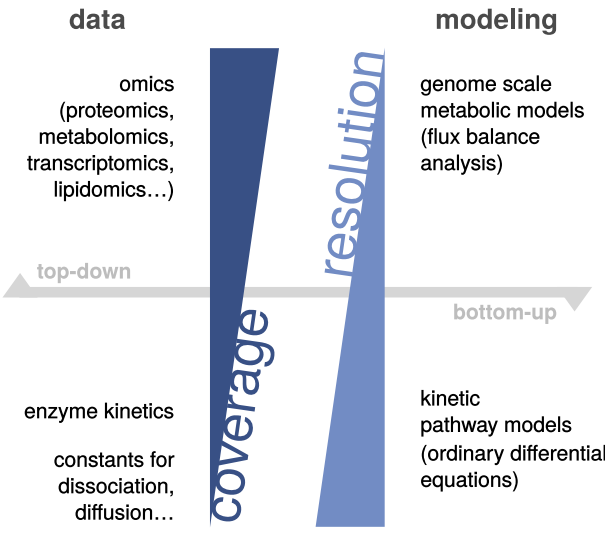


Figure 1.3: Coverage and resolution of systems biological approaches, with focus on the terms relevant for this work.

As models will always include reductions, simplifications and approximations, their level of granularity has to be chosen with care, consistent with the corresponding data (Figure 1.3). Accordingly, models are as good as their data in terms of resolution and accuracy. However, non-intuitive, mechanistic conclusions can be drawn from them and, moreover, consistency of the data and its compliance to theoretical concepts can be tested rigorously. From consistent models, the system's behavior in other contexts can be predicted, including scenarios that are inaccessible to experimental investigation.
The more accurate models are and the more dependencies they include, the more difficult it is to analyze and simulate them. Omics-size models are therefore often simplified, such as purely stoichiometric models of the entire metabolic network of a cell, or face problems of combinatorial expansion or limiting computational capacity. [93, 141]

Need for specific questions

In the light of these restrictions, modeling efforts require specific questions. Based on the nature of the desired knowledge, the model resolution needs to be adapted and matching data has to be included or requested. The model can then either give a direct answer to the posed question or propose a set of testable hypotheses, complying with the current state of knowledge and data. In the best of all worlds, experimental and modeling efforts can complement each other and iteratively advance understanding of the biological system.

1.4 A set of questions

How is the central carbon metabolism embedded in the cellular metabolic network?
Which of its reactions are thermodynamically limiting and are they limiting in all conditions?
Where are regulatory points in the mycobacterial central carbon metabolism?
Which concepts facilitate metabolic flexibility in mycobacteria?

In this thesis, the above questions are addressed by integrating mathematical modeling and experimental data to study the central carbon metabolism in mycobacteria. The following Section outlines the contents of this work, a graphical overview is given in Figure 1.4.
The analysis starts from an examination of the available data for modeling the central carbon metabolism in mycobacteria. For the purpose of this work, a set of experiments was carried out by experimental partners in which proteomics and metabolomics data were acquired, as is described in detail in Chapter 3. Chapter 4 consists of a careful analysis of these data sets to highlight the insight that can be attained from the data already without any modeling involved. The processed data are the basis for all further results in the thesis.
The subsequent Chapters will employ different modeling approaches to arrive at a deeper understanding of the mechanistic processes underlying the data. The work thereby zooms in from a network based view of the entire central carbon metabolism to a detailed, dynamic description of individual, pathogenicity related pathways.
By formalizing a method to systematically analyze and quantify boundary fluxes (biomass backtracking) of the central metabolic system, the demand that remote biosynthetic processes exert in order to supply functional biomass for bacterial growth is charted (Chapter 5.1). Based on these boundary fluxes, the thermodynamic landscape of the system will be characterized. A linear programming approach is used to combine and extend existing methods such as the Maximum-Minimum driving force and thermodynamic flux balance analysis to the network

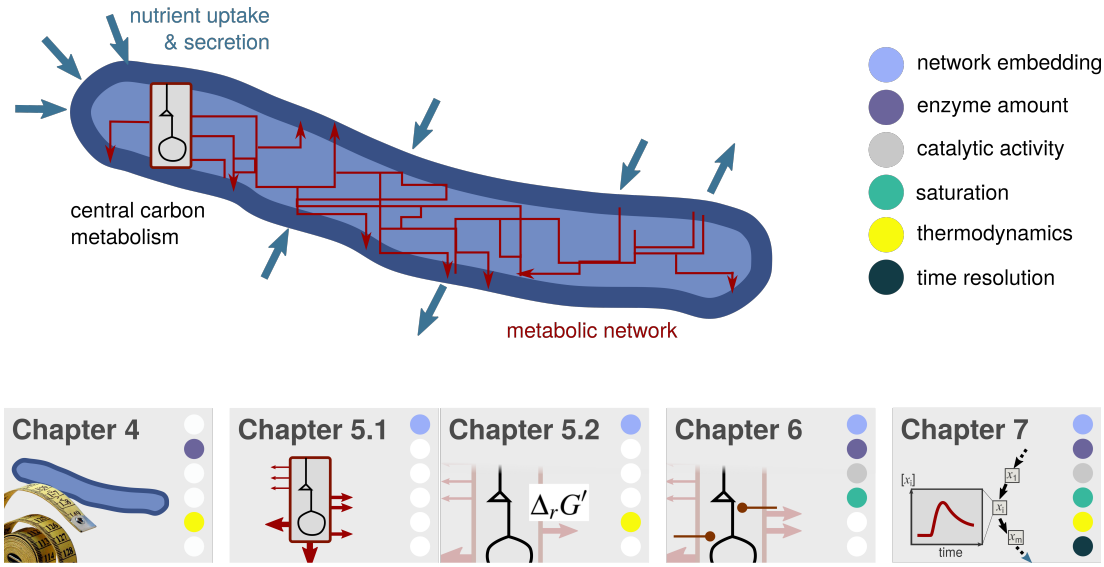


Figure 1.4: Overview of the results of this work, for a detailed description please refer to the text.

context, while integrating absolute measurements of metabolite concentrations (Chapter 5.2). In Chapter 6, the stoichiometric and thermodynamic analysis will be complemented by enzyme kinetic equations. A consistent, literature derived model is presented, covering a large part of the current knowledge on the central carbon metabolism of *M. tuberculosis* (Chapter 6.1). The model is then used to exemplify a novel method to predict regulatory sites in metabolic networks as a minimum set of interventions which guarantee data consistency in steady state (Chapter 6.2), focusing on the transition between experimental carbon sources. The quantitative metabolomics and proteomics data are at the core of this analysis.
Leaving the realms of steady state, Chapter 7 further zooms in onto a specific metabolic function of *M. tuberculosis*, the catabolism of the potential *in vivo* carbon source propionyl-CoA. The role of two seemingly redundant pathways for the robust execution of this task is resolved by means of thermodynamic-kinetic modeling. Accordingly, specific functions can be assigned to each of the two pathways: While one acts as a transcriptionally regulated, high-capacity catabolizer, the other can dynamically detoxify sudden catabolic challenges. In the pathological context, the results explain one feature of flexibility of the mycobacterial metabolism.

II Methods

2. Computational methods

This Chapter summarizes all computational methods employed within this work. Methods and techniques for processing and analysis of the experimental data are introduced in short. The main focus lies on the description of the methods required for the computational modeling central to this work. General approaches of metabolic modeling are explained, introducing concepts of flux balancing, kinetic modeling as well as biological thermodynamics. The required numerical methods to simulate and optimize such models are be described subsequently.

2.1 Statistics, data analysis and integration

2.1.1 Integration of various data types

The modeling approaches presented in the following integrate data from various data sources, types of experiments and measurements as summarized in Figure 2.1. The integration requires standardized and careful unit conversions as well as consideration of the cell state and environmental conditions that the data were derived from, as will be described later. In all cases, quantitative data are preferred, whereby also relative data in form of fold changes to a reference condition can be included via appropriate observables (*cf.* Chapter 2.3.4).

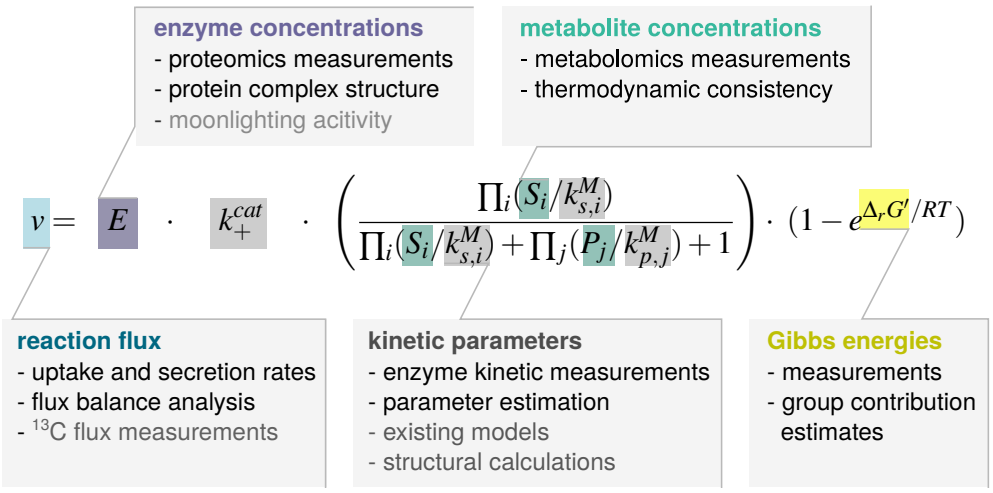


Figure 2.1: Possible ways to integrate data into kinetic network models on the example of a thermodynamic form of the Michaelis-Menten kinetic [136]. All factors in black were applied within the scope of this work, gray denotes possible other factors.

Table 2.1: List of all used databases and their purposes.

Database	Used for	Reference
KEGG	Referencing compounds and reactions, reaction definition	(84)
ChEBI	Referencing compounds	(69)
Tuberculist	*M. tuberculosis* gene database, gene names, functions	(101)
BRENDA	Enzyme kinetic parameters, regulators	(175)
Pfam	Detection of regulated protein domains	(55)
UniProt	Identifier mapping between databases, protein referencing	(12)
Bionumbers	General biological numbers	(123)
eQuilibrator	Gibbs energies, consistency of reactions	(58)

2.1.2 Database mining

Information on the central carbon metabolism in mycobacteria was collected from several databases. Where available, application programming interfaces (API) were used to extract data. The used databases are summarized in Table 2.1.

2.1.3 Classification and enrichment

Principal component analysis

Principle component analysis (PCA, [143]) was used to classify the experimental data and to identify general trends and groups of behaviors. The analysis decomposes a dataset, commonly in the form of a matrix, into the major components able to explain the observed total variance. In multivariate statistics, PCA is used to identify groups of variables with a common behavior. The analysis is based on the eigenvalue decomposition of the covariance matrix of the data. Here, the Matlab implementation in the function `pca` was used.

Differential expression and enrichment tests

Experimental proteomics data were assumed to be log-normally distributed. Proteins were classified as differentially expressed, when the null-hypothesis that the logarithmic measurements of the current condition have the same mean expression like all other conditions, was rejected with a p-value smaller than 0.01 in a two sided t-test. Enrichment tests were conducted using the statistical over-representation test in the PANTHER online interface [121], with the whole *M. tuberculosis* genome as background.

2.2 Modeling methods

Metabolic modeling intends to describe the behavior of cellular metabolites. Their intracellular concentrations $\mathbf{x}$ are altered with a given stoichiometry by the action of enzymatic reactions carrying out chemical transformations or by source and sink reactions across the boundaries of the modeled system. The fluxes through all reactions, *i.e.* the amount of metabolites processed per unit of time, can be summarized in the flux vector $\mathbf{v}$. The matrix of the linking stoichiometric coefficients $\vartheta_{i,j}$ is called ***stoichiometric matrix*** $\mathbf{N}$. The system of metabolites and reactions can then be expressed mathematically as

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{N} \cdot \mathbf{v}. \tag{2.1}$$

In the following, used methods to simulate and optimize such systems are described.

2.2.1 Flux balance analysis on the genome scale

Genome-scale metabolic reconstructions

Genome-scale metabolic reconstructions collect all information on the metabolic network of a species: Known reactions, linked to their catalyzing enzymes; all involved metabolites, potentially associated with their localization in a specific compartment; uptake and secretion reactions, potentially via their specific transporters. The reconstruction is usually based on genomic information as well as literature and database mining. Numerous methods exist to curate the reconstructions, such as mass and charge balancing or gap filling. The reconstructions usually include an artificial reaction that consumes metabolites towards the formation of biomass, with stoichiometries based on experimental assessment of the cellular composition. Also the cost of maintenance, payed usually in form of ATP, is included in the reconstruction. In that way, the genome-scale metabolic reconstruction provides the stoichiometric matrix $\mathbf{N}$ of the whole known metabolism of a cell in a well annotated manner, which can now be used for modeling. [52, 192]

Flux balance analysis

By assuming the metabolic system to be in steady state ($\mathbf{N} \cdot \bar{\mathbf{v}} = 0$), ***flux balance analysis*** (FBA, [53]) can calculate a distribution of fluxes in the network. This distribution is shaped by the constraints imposed by the network topology, encoded in $\mathbf{N}$, as well as by the chosen boundary fluxes (*i.e.* allowed uptakes and secretion routes), which can be bounded by experimental measurements. The constraints thereby define a whole space of feasible flux distributions.
To transition from feasibility to biologically relevant optimality, objective functions $\mathbf{c}$ that mimic possible cellular objectives can be formulated. These objectives are usually linear combinations of fluxes describing for example the maximization of biomass, an overall low total flux associated with a low enzyme burden, maximization of ATP production or the maximization of the production of a desired chemical compound in metabolic engineering [178].
Mathematically, FBA can be described as a linear, constrained optimization problem

$$\begin{aligned} \max_{v} \quad & \mathbf{c}^{\mathrm{T}} \cdot \mathbf{v} \\ \text{subject to} \quad & \mathbf{N} \cdot \mathbf{v} = 0 \\ & \mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub}, \end{aligned} \tag{2.2}$$

with $\mathbf{v}_{lb}$ and $\mathbf{v}_{ub}$ as lower and upper bounds on the fluxes.
FBA has proven to be useful in a wide range of biological and biotechnological applications and a whole genealogy of extending methods is available to integrate additional constraints and types of data [23]. On the downside, FBA alone does not process metabolite concentration values and their impact on the fluxes. Also the lacking time resolution renders it insufficient for many common applications.

2.2.2 Kinetic modeling

Ordinary differential equations

While genome-scale, constraint-based methods focus on the cellular metabolism in steady state, much information on biological functionality is encoded on the transient time-scale. Describing functional kinetic mechanisms of enzyme action is therefore bound to add a more detailed level of understanding to the static descriptions.

Ordinary differential equations (ODEs) are commonly used to model cellular processes on a mechanistic level and have proven to be useful tools in many studies [18, 31, 94, 190]. In contrast to the FBA formulation, the metabolic system's time evolution is here described by a set of ODEs for the metabolite concentrations $\mathbf{x}$ rater than a set of linear equations employing only the numeric value of the flux $\bar{\mathbf{v}}$ in steady state. Thereby, each ODE formulates as the sum of all fluxes producing or consuming metabolite x_i with the stoichiometric coefficients in $\mathbf{N}$:

$$\frac{\partial x_i}{\partial t} = \sum_{j=1}^{n_r} \vartheta_{i,j} \cdot v_j(E_j, \mathbf{x}, \boldsymbol{\theta}), \tag{2.3}$$

with n_r as the total number of reactions in the system. The fluxes $\mathbf{v}$ are thereby functions of the enzyme concentrations $\mathbf{E}$, the metabolite levels $\mathbf{x}$ as well as of a set of parameters $\boldsymbol{\theta}$, which can include enzyme kinetic constants, thermodynamic parameters, activation or inhibition constants and others.
The ODE formulation relies on a number of assumptions that have to be examined critically for their validity for each new model. A well stirred system is assumed, where no spatial concentration gradients exist. Intracellular concentrations are also assumed to be normally distributed, for example across a cell population, such that they can be described reliably by their mean. While capturing time evolutions and mechanistic dependencies of the fluxes on metabolite concentrations, ODE systems are often large and complex with a vast number of potentially unknown parameters.

Reaction kinetics

A review of available kinetic laws and their parameterization in the -omics era was published within the scope of this work as: Tummler, K., Lubitz, T., Schelker, M. & Klipp, E. (2013) New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. FEBS journal 281(2):549-71

The enzymatic mechanism is modeled by a specific rate law for each flux v_j, describing the dependencies of the reaction rate on the concentrations of substrates, products and the catalyzing enzyme (*cf.* Equation 1.1 in the Introduction). The rate laws can also include influences of activating or inhibiting compounds, thermodynamic relations, effects of post-translational modifications and other factors. In their most general form, the rate laws constitute of an enzyme concentration dependent term f and an independent term g which are connected multiplicatively:

$$v_j = f(E_j) \cdot g(\mathbf{x}, \boldsymbol{\theta}) \tag{2.4}$$

Hence, the flux will linearly depend on the (active) enzyme concentration, given all other variables are held constant. The enzyme concentration is thereby equal to the concentration of active catalytic sites for the specific reaction, considering enzyme complex structure and possible moonlighting activities [46, 81]. The function g is used to describe the molecular mechanism of the catalyzing enzyme. While the rate of uncatalyzed chemical reactions only depends linearly on the concentration of the substrates according to the law of mass action, enzyme catalyzed reactions usually show saturating behavior for high substrate concentrations,

coinciding with all available enzyme molecules being occupied. The first and very famous model of this observation was formulated in 1913 as the ***Michaelis-Menten kinetic*** [122]:

$$v_j = v^{max} \cdot \frac{S}{S + k^M} \tag{2.5}$$

with S as substrate concentration, v^{max} as maximum possible reaction rate, given by the product of enzyme concentration E and the catalytic activity k^{cat}, and k^M as the Michaelis-Menten constant, signifying the substrate concentration at which the rate reaches half its maximum. According to the above notation, the parameter vector $\boldsymbol{\theta}$ for a Michaelis-Menten flux would contain entries for v^{max} and k^M. The kinetic law is derived from the full set of mass action equations describing binding, catalysis and dissociation of enzyme, substrate and product, and assuming a fast equilibration of the substrate binding reaction. Similarly, more detailed rate laws have been derived over the years to describe the action of several substrates, binding order, interactions with activating or inhibiting compounds, competition, cooperativity and other effects [196]. These rate laws provide a higher level of mechanistic accuracy in the description of enzyme action at the price of extremely complex mathematical terms with a large number of kinetic parameters.

Metabolic control analysis

Metabolic control analysis (MCA) provides a well established framework to quantify the control of single reactions or species on the flux of a metabolic network [70, 83]. Despite being limited to characterizing systems in steady state, MCA gives insight into the systemic properties of metabolic pathways and the interdependencies of the participating agents. In practice, MCA requires a fully parameterized dynamic network model for the calculation of the control distribution, which is not always available, especially for larger systems.

MCA distinguishes between ***control coefficients***, which measure the effect of small changes in the participating enzymes' concentrations $\mathbf{E}$ on redistributing (*i.e.* controlling) the steady state flux $\bar{\mathbf{v}}$ or the steady state concentrations $\bar{\mathbf{x}}$ of the entire system, and ***response coefficients***, which quantify the response of steady state fluxes $\bar{\mathbf{v}}$ and concentrations $\bar{\mathbf{x}}$ to small perturbations in the kinetic parameters θ. Flux control coefficients $C_{v_i}^{\bar{v}_j}$ are defined as

$$C_{v_i}^{\bar{v}_j} = \frac{v_i}{\bar{v}_j} \frac{\partial \bar{v}_j}{\partial v_i} \tag{2.6}$$

and signify the effect of small perturbations in a specific flux v_i on flux $\bar{v}_j$ in the new steady state. Please note, that $\bar{\mathbf{v}}$ is used as the vector containing all network fluxes in steady state for consistency with the above definitions. In the original MCA works this entity is usually referred to as J. The control coefficients in the above equations are the normalized ones. Biologically, the perturbations can be due to a change in the total enzyme concentration or caused by enzyme activations and inhibitions. The control coefficients differ from classical sensitivities, as they describe the final effect after the equilibration to a new steady state following the perturbation rather than immediate responses. Concentration control coefficients are defined accordingly as

$$C_{v_i}^{\bar{x}_j} = \frac{v_i}{\bar{x}_j} \frac{\partial \bar{x}_j}{\partial v_i}, \tag{2.7}$$

with $\bar{\mathbf{x}}$ as the vector of metabolite concentrations in steady state. Flux and concentration response coefficients are analogously composed of partial derivatives of $\bar{\mathbf{v}}$ and $\bar{\mathbf{x}}$ with respect to all kinetic parameters in $\boldsymbol{\theta}$.

In the scope of this work, a version of metabolic control analysis was implemented in Matlab, to be used in the D2D framework [156]. The calculation is based on the matrix notation of the control and response coefficients (described in detail *e.g.* in [72]).

2.2.3 Thermodynamic methods

Besides the dependency of the flux on the concentrations of the reaction's educts, also the chemical nature of the transformation restricts the reaction rate. The following Section summarizes how these effects can be included into metabolic network modeling with the help of biological thermodynamics.

Thermodynamic entities and their relation

The direction, in which a reaction runs, is defined by the energetics of the chemical transformation occurring. If energy is released, the reaction can happen spontaneously according to the second law of thermodynamics. The amount of energy released is measured as change in the system's ***Gibbs free energy*** G in kJ/mol. If the change in free energy due to the reaction, $\Delta_r G'$, is negative, energy is released. The reaction is feasible in this direction and called exergonic. If $\Delta_r G'$ is positive for a reaction, it will be infeasible and is called endergonic. At the same time, the same reaction in reverse direction will be feasible, with a free energy change of -$\Delta_r G'$.
Feasible thereby means that the flux in one reaction direction is larger than in the reverse one, not that the flux in the reverse direction vanishes. Accordingly, $\Delta_r G'$ defines the ratio between forward J^+ and backward J^- flux through the reaction in the ***flux force relationship*** [16] as

$$J^- = e^{\Delta_r G'/RT} \cdot J^+, \tag{2.8}$$

with R as the universal gas constant in kJ/K and T as the absolute temperature in K. The net-flux v_j, which is observed macroscopically, is thereby the difference between J_j^+ and J_j^-.
At $\Delta_r G' = 0$, J^+ equals J^- and the reaction is in equilibrium. Again, this does not refer to a stalling reaction, but rather that the fluxes in forward and backward direction are exactly the same, such that no net-flux occurs (*i.e.* no substrate or product is effectively produced). The concentrations of substrates and products at equilibrium define the ***equilibrium constant*** k_{eq} as

$$k_{eq} = \prod_{i=1}^{n_x} \bar{x}_i^{\vartheta_i}, \tag{2.9}$$

with $\bar{\mathbf{x}}$ as the (active) concentration of the educts of the reaction at steady state, ϑ as their corresponding stoichiometric coefficients and n_x as the number of metabolites participating in the reaction. The product in the above equation is often referred to as the reaction quotient Q or as mass action ratio Γ.
The equilibrium constant is an intrinsic property of a chemical reaction. It is linked to the ***standard Gibbs free energy*** $\Delta_r G^0$ of a reaction by

$$\Delta_r G^0 = -RT \cdot \ln\, k_{eq}. \tag{2.10}$$

Equilibrium constants and hence $\Delta_r G^0$ can be assessed experimentally by measuring substrate and product concentration in an equilibrated solution [66]. As equilibration times for many reactions are extremely long, it is also useful to calculate $\Delta_r G^0$ from the ***standard Gibbs formation energies*** $\Delta_f G^0$ of the participating educts, signifying the amount of energy required to form them from their elements in standard state, as

$$\Delta_r G^0 = \sum_{i=1}^{n_x} \vartheta_i \cdot \Delta_f G^0. \tag{2.11}$$

$\Delta_f G^0$can be measured experimentally or estimated by group contribution [117, 136]. Both k_{eq} and $\Delta_r G^0$ can be used to calculate the current thermodynamic potential $\Delta_r G'$ of a reaction, which additionally depends on the concentrations of its substrates and products. The reaction quotient Q, as already defined in equilibrium for k_{eq}, thereby quantifies how much the potential is shifted relative to $\Delta_r G^0$ due to concentrations differing from their equilibrium value. If, for example, substrate concentrations are drastically decreased, the thermodynamic gradient will become shallower in the reaction direction that consumes the substrate. The concentration adjusted $\Delta_r G'$ calculates as

$$\Delta_r G' = \Delta_r G^0 + RT \cdot ln\ Q. \tag{2.12}$$

Using measured intracellular metabolite concentrations, $\Delta_r G'$ can then give insight into the thermodynamic landscape that constrains cellular metabolism. [3]

Assessing reaction reversibilities

To quantify how far a reaction is from its equilibrium given a set of educt concentrations, the ***disequilibrium ratio*** ρ can be used. It is defined as

$$\rho = Q/k_{eq}. \tag{2.13}$$

If ρ is smaller than 1, the reaction is feasible in the forward direction, for ρ larger than 1 in the backward direction. ρ can only be calculated if all educt concentrations are known. Additionally, especially when cellular metabolic regulation is investigated, it is interesting how much effort, in terms of changes in metabolite concentrations, it would require to change the direction of the reaction. For FBA approaches, a sound definition of the reactions' potential to invert their direction is considered essential. However, this potential for reversal also depends on the stoichiometry of a reaction: It is more effort for the cell to reverse a 1:1 reaction than, *e.g.* one with 3:3 stoichiometry, as the required change in Q has to be achieved by adjusting only one concentration on either side of the reaction as opposed to the product of three concentrations.
To assess these issues of uncertain or unknown concentrations and variable stoichiometry, the ***reversibility index*** $\hat{\gamma}$ can be used as a stoichiometry adjusted measure for how reversible a reaction is [133]. $\hat{\gamma}$ assumes all educt concentrations to lie within a logarithmically symmetric range around a sensitive physiological concentration C (usually 100 μM). With n_s as the number of substrates and n_p as the number of products, the reaction quotient then reads as $Q_{RI} = C^{n_p - n_s}$. With this, the reversibility index is defined as

$$\hat{\gamma} = \left(\frac{k_{eq}}{Q_{RI}} \right)^{2/(n_p + n_s)} \tag{2.14}$$

For quantitative assessment, the distance of $\hat{\gamma}$ to 1 can be used. The larger it is, the more irreversible is the reaction. Noor *et al.* [133] used a threshold of $\hat{\gamma} > 10^3$ to classify a reaction as irreversible in the forward direction (accordingly $\hat{\gamma} < 10^{-3}$ in backward direction). If the concentration of a metabolite is known, it can be substituted for the default concentration C in Q_{RI}, thereby reducing the number of unknown substrates n_S or products n_p by one.

Integration of thermodynamic constraints in kinetic reaction equations

For biological systems, the energetic restriction of metabolic reactions gives rise to several constraints on the feasible space for kinetic parameters. The ***Haldane relationship*** [25] formalizes the dependency of Michaelis-Menten constants and catalytic activities on the equilibrium constant:

$$k_{eq} = \frac{k_+^{cat}/k_s^M}{k_-^{cat}/k_p^M}. \tag{2.15}$$

For all parameter sets $\boldsymbol{\theta}$ fulfilling this relationship, the net-flux through the corresponding reaction will vanish when the species reach their equilibrium concentrations as defined by k_{eq}. For systems of reactions that form a loop, all associated equilibrium constants have to fulfill the ***Wegscheider condition***

$$\prod_{i=1}^{n_l} k_{eq,i} = 1, \tag{2.16}$$

with n_l being the number of reactions involved in the loop.
In works of thermodynamic FBA this is also referred to as a generalization of Kirchoff's loop law for electrical circuits [14, 147]. More generally, both approaches state that the conservation of energy forbids circular fluxes (currents) to occur.
Several works have proposed kinetic rate laws that integrate these thermodynamic constraints. The rate of an reaction is not completely defined by its thermodynamic gradient, *i.e.* thermodynamically feasible reactions can still take very long to equilibrate. However, the flux force ratio which is defined by the gradient (Equation 2.8) will fix the extent of the net flux and should hence be considered for kinetic modeling of metabolic pathways.
The thermodynamic version of the ***convenience kinetic***, introduced by Liebermeister and Klipp [104], is one example of such rate laws. It is convenient, because reactions of arbitrary stoichiometry can be described in an automated manner. Similarly, the ***modular rate laws*** [106] allow the integration of regulatory terms alongside the thermodynamically consistent kinetic formulation. ***Thermodynamic kinetic modeling*** (TKM, [50]) was developed in analogy to electrical engineering and can be applied to transform kinetic laws to their thermodynamically consistent form. Noor *et al.* [136] proposed a decomposed rate law derived from the Michaelis-Menten kinetic that has the benefit of a clear separation of the kinetic and thermodynamic contributions to the reaction speed, which can then be analyzed individually.

2.3 Numeric methods

2.3.1 Numeric integration of ODE systems

If each reaction in the entire or a considered subpart of a cellular metabolic network is described by a more or less complex enzymatic rate law, the resulting system of differential equations is far beyond analytic solvability. Such systems are commonly solved by numeric integration, for which a number of computational algorithms are available. Mathematically, the problem can be formulated as an initial value problem of approximating a function $y(t)$ that is the solution of

$$\frac{\partial y(t)}{\partial t} = f(t, y(t), \theta) \tag{2.17}$$
$$\text{with} \quad y(t_0) = y_0.$$

Please note that in the following the notation $y(t)$ for a single variable is used, but the methods are directly applicable to sets of ODEs of several variables $\mathbf{y}(t)$. Numeric methods try to subsequently extrapolate values of y for the next time step t_{n+1} using the value at the current time $y(t_n)$ (explicit methods), from previous time steps $y(t_{n-1\cdots i})$ (multi step methods) or the next time step $y(t_{n+1})$ (implicit methods) and their derivatives (*cf.* Table 2.2). In implicit methods the value of $y(t_{n+1})$ is required to solve for $y(t_{n+1})$, resulting in an usually nonlinear equation, which can for example be solved using Newton's method (see Chapter 2.3.3). This extra computational cost is redeemed by a higher accuracy of the solution.
Originating from the explicit Euler method, which walks along the tangent at the current value $y(t_n)$ for a small time step h, sets of more complex methods have been developed to reduce the prediction error. Runge-Kutta methods calculate and average additional increments k_i at different positions within the h-step, to correct for over- or underestimation of the change by the derivative at the current point. The number of increments used defines the order of the method, with more increments improving accuracy but complicating the calculation. Adams-Bashforth type methods take into account the calculated function and derivative values of the previous time steps to increase accuracy. Also for them, the number of previous time steps used defines the order of the method and a higher number results in higher accuracy. Adams-Moulton and Backwards differentiation formulas (BDF) are extensions of the implicit backward Euler method in the same manner, each with their specific way of defining the coefficients for averaging.
The smaller the step size h is, the more accurate the solution for all mentioned methods will be, at the cost of increasing calculation times. Around time points, where fast changes in the value of y can be observed, this step size reduction is most beneficial. Accordingly, most modern algorithms employ an adaptive regulation of the step size, relative to the steepness of f, to increase accuracy while limiting computational costs. If the order of the methods is also changed accordingly, the algorithms are referred to as variable step-size, variable order (VSVO).
Especially when simulating biological systems, such changes in the steepness of the derivative do not only occur at different times but also for different species - a problem referred to as ***stiffness***. Stiff equation systems lead to issues of stability in the numerical algorithms and force extremely small step sizes and long computation times, with results being error prone. Stiffness issues are commonly addressed with BDM and Adams-Moulton type algorithms. [27, 146]

Table 2.2: Selection of common methods for numeric integration, along with their calculation rules. *s will denote the order of the Runge-Kutta method, ** shown is the 2-step Adams-Bashforth method

Method	**Approximation**	**Type**
Euler method	$y(t_{n+1}) = y(t_n) + h \cdot f(t_n, y(t_n), \theta)$	explicit
Runge-Kutta*	$y(t_{n+1}) = y(t_n) + h\sum_{i=1}^{s} b_i \cdot k_i$	explicit
Backward Euler	$y(t_{n+1}) = y(t_n) + h \cdot f(t_{n+1}, y(t_{n+1}), \theta)$	implicit
Trapezoidal rule	$y(t_{n+1}) = y(t_n) + \frac{1}{2}h\,(f(t_{n+1}, y(t_{n+1}), \theta) + f(t_n, y(t_n), \theta))$	implicit
Adams-Bashforth**	$y(t_{n+2}) = y(t_{n+1}) + \frac{3}{2}hf(t_{n+1}, y(t_{n+1}), \theta) - \frac{1}{2}hf(t_n, y(t_n), \theta)$	implicit, multi-step

Simulation standards

Metabolite trajectories $\mathbf{x}(t)$ were simulated in logarithmic space in μM concentrations. A fixed conversion factor was calculated based on cell size and the number of cells per optical density (OD) unit to convert from the experimentally measured concentrations in pmol/mL$_{\mathrm{OD}_{600}=1}$ to intracellular concentrations as described in Section 3.1.

2.3.2 Optimization problems

The information collected from literature and biology during the construction of a model does in the vast majority of cases not define its behavior uniquely. The open question remains, which behavior is the best or most realistic one in the biological context. Optimization methods tackle this problem by systematically tuning the model's degrees of freedom, such as unknown parameters or boundary fluxes, to achieve a behavior best matching experimental data. In metabolic modeling, many optimization approaches have been employed to learn about biological functionality. They mainly fall into two groups: (i) FBA based approaches employ linear optimization to calculate system wide flux distributions and (ii) kinetic models rely on nonlinear optimization for the estimation of model parameters to reproduce the time evolution of metabolic entities. In the following Section, both aspects will be explained as far as they were applied in this work.

Linear optimization problems

In its standard form, a linear optimization problem describes a set of linear constraints that bounds the feasible solution space for the variables $\mathbf{d}$. Please note that the vector of variables $\mathbf{d}$ can contain any type of entity such as concentrations x, fluxes v, Gibbs energies $\Delta_r G'$ and others. The constraints can define equality or inequality relations. Often an objective function $\mathbf{c}$ as a linear combination of values in $\mathbf{d}$ is defined, which is intended to be optimized. The standard form of the problem then reads as:

$$\begin{aligned} &\max_{\mathbf{d}} \ \mathbf{c}^{\mathrm{T}}\mathbf{d} \\ &\text{subject to } \ \mathbf{A}\cdot\mathbf{d} \leq \mathbf{b}. \end{aligned} \tag{2.18}$$

Depending on the rank of the matrix $[\mathbf{A}\ \mathbf{b}]$, which collects all coefficients of the equality and inequality relations, the system can have no, one or infinitely many solutions. Please note that the equation shows an inequality constraint, but equalities can be included in the same manner.

Duality

Each linear optimization problem has a corresponding dual problem of the form

$$\begin{aligned} &\min_{\mathbf{l}} \ \mathbf{b}^{\mathrm{T}}\mathbf{l} \\ &\text{subject to } \ \mathbf{l}^{\mathrm{T}}\cdot\mathbf{A} = \mathbf{c}, \ \ \mathbf{l} \geq 0. \end{aligned} \tag{2.19}$$

Similarly, equality constraints can be reformulated as inequalities ($\leq$) in the dual problem. If the original (primal) problem has a unique solution $\mathbf{d}^*$, it will define the solution $\mathbf{l}^*$ to the dual problem as $\mathbf{c}^{\mathrm{T}}\mathbf{d}^* = \mathbf{b}^{\mathrm{T}}\mathbf{l}^*$.
For each constraint in the primal problem, one variable exists in the dual problem. The dual variables $\mathbf{l}$ are also referred to as ***shadow prices***. Conceptually, a shadow price describes how much the optimum value of objective function $\mathbf{c}^{\mathrm{T}}\mathbf{d}^*$ would improve if the corresponding constraint would be relaxed by one unit. Shadow prices can therefore be used to investigate where constraints are exhausted and hint to effectively restricting variables and constraints. [119, 134, 146]

Nonlinear Optimization Problems

If a system of equations contains any kind of nonlinear terms with respect to the variables $\mathbf{d}$, its optimization becomes more challenging. A general nonlinear optimization can be formalized as

$$\max_{\mathbf{d}} g(\mathbf{d}, \Theta) \tag{2.20}$$

Where $g(\mathbf{d}, \Theta)$ is a nonlinear function of the systems variables and possibly of external parameters Θ. The system can also be subjected to constraints on $\mathbf{d}$, which can be linear or nonlinear equality or inequality constraints.

2.3.3 Numeric optimization algorithms

Linear Optimization Algorithms

Linear optimization problems are comparably easy to solve, a number of efficient algorithms are available. The most commonly used algorithm is the ***Simplex algorithm*** [45]. The algorithm exploits the fact that the feasible solution space is a convex, possibly unbounded polytope in the variable space. If an optimum of the objective function $\mathbf{c}^{\mathrm{T}}\mathbf{d}$ exists, it will lie on an edge or vertex of that polytope. The algorithm first finds an initial extreme point on the hull of the polytope (*i.e.* a corner) and then follows the edges in the direction of improving objective function to the next extreme point. It thereby visits points with better objective values until the next step does not yield any further improvement or an unbounded edge is reached (termination criteria). The initial point can be found by a number of different methods, however, if no initial feasible point can be found, the feasible solution space is empty and the problem has no solution. The algorithm has been improved steadily since its development, resulting in a faster convergence. Even though its worst case complexity is exponential, meaning that the time to compute a solution depends on the number of variables in an exponential manner, for the vast majority of applications the algorithm converges faster (average case complexity, polynomial).

Interior point algorithms advance through the interior of the polytope using *e.g.* Newton's method. The system's constraints are thereby encoded in logarithmic barrier functions, directly added to the objective. Interior point algorithms can also make use of the dual problem to improve convergence. They have polynomial worst case convergence, but are in practice often slower than simplex methods. [146]

For many practical linear optimization problems, the variables $\mathbf{d}$ can only take discrete or boolean values. Such ***mixed integer linear programming*** problems (MILP) can be solved with ***branch-and-cut/bound algorithms*** (BC/BB) that approximate the integer solution by solving the corresponding continuous problem (*e.g.* with the simplex algorithm) and adding new constraints if the solution is not integer-valued. The solution of MILP problems is computationally difficult and can have very long convergence times in practice.

Nonlinear optimization algorithms

The feasible solution space of nonlinear problems does not necessarily have the shape of a convex polytope and hence its beneficial characteristics cannot be exploited by algorithms. Most nonlinear optimization algorithms apply iterative approaches to move in the allowed variable space towards regions of lower objective values. Some commonly used algorithms, part of which were also applied in this work, are described briefly here.

The ***Nelder-Mead*** version of the simplex algorithm [131] tries to search the variable space based only on evaluation of the objective function g. It spans a polytope in the variable space with one additional dimension (*e.g.* a triangle in a search space of two variables) and moves towards the

vertex associated with the best objective value by a series of transformation events (flipping, contraction, expansion). In practice, such so called ***derivative free*** methods can usually be improved by considering the derivative of the objective function with regard to the variables.
Derivative based algorithms generally try to find the direction in the search space in which the decrease in the objective function is steepest. Based on this steepness, the algorithms estimate how far they can likely step in that direction to maximally improve the objective value. Many different methods are available to calculate or approximate the descent and the step size of an iteration step $\mathbf{h}$. ***Newton's method***, as a method to find roots of functions, is used in optimization to find the root of the first derivative of g, *i.e.* the extremum of g. For example in an one dimensional search space, the step $\mathbf{h}$ taken from the current point $\mathbf{d}_n$ in the variable space is calculated as

$$\mathbf{h} = -\gamma \frac{\partial g(\mathbf{d}_n)}{\partial \mathbf{d}} \Bigg/ \frac{\partial^2 g(\mathbf{d}_n)}{\partial^2 \mathbf{d}} , \tag{2.21}$$

based on a Taylor expansion of g. The scaling parameter γ can be used to adjust the step length. The objective value $g(\mathbf{d}_n + \mathbf{h})$ is then hopefully, and most likely, better than the current one $g(\mathbf{d}_n)$. In higher dimensions, the first and second derivatives, also referred to as the Jacobian matrix $\mathbf{J}$ and the Hessian matrix $\mathbf{H}$, respectively, are used to calculate the direction of the steepest descent in the variable space. The usage of the Hessian improves convergence compared to calculating h based only on the Jacobian, as is done for example in the ***gradient descent*** method. However, often the calculation of Hessians is computationally expensive and sometimes numerically error-prone. Both methods from above also show poor convergence in shallow regions of the variable space. To resolve these problems, more sophisticated methods combine the Newtonian root finding with other numeric steps to improve convergence. ***Trust region*** algorithms approximate g within a neighborhood – the trust region – of $\mathbf{d}_n$ by a simple, often quadratic function that can be easily solved. $\mathbf{h}$ is then calculated based on the approximation. If $g(\mathbf{d}_n + \mathbf{h})$ is a better point the trust region shrinks, until it is small enough for the desired accuracy of the solution. In contrast, ***line search*** methods take the Newtonian direction of the step, but adapt its length (γ in Equation 2.21) at every iteration step to approach the optimum faster, but also to avoid overshooting. For constraint problems, ***interior point*** and ***active set*** algorithms can be used. They implement penalty terms for violating constraints, which have to be minimized alongside the objective function.
Often objective functions of optimization problems take the form of sums of squared differences, so-called residuals $\mathbf{r}$, as for example when distances of a model to several measured data points are to be minimized. Algorithms specifically and efficiently dealing with such problems are the ***Gauss-Newton*** method that requires only the Jacobian to calculate the step $\mathbf{h}$ as

$$h = -\frac{\mathbf{J}_r^{\mathrm{T}} \mathbf{r}(\mathbf{d}_n)}{\mathbf{J}_r^{\mathrm{T}} \mathbf{J}_r}, \tag{2.22}$$

with $\mathbf{J}_r$ as the Jacobian of the residuals with respect to the variables $\mathbf{d}$. The ***Levenberg-Marquardt*** [100, 113] algorithm interpolates between Gauss-Newton and gradient descent steps to improve convergence. [146]
The selection of the appropriate algorithm depends on the specific problem as well as the available computational capacity. Often it is helpful to test the performance of several algorithms in the specific context.

Global optimization heuristics

For linear optimization problems, if they are feasible, the found optimum will contain the best function value in the entire variable space – the global optimum. Non-linear problems, on the other hand, can contain multiple, local optima of which only one will be the best. The algorithms described above will be trapped in a local optimum, due to their dependence on the gradients which are positive in any direction from the local optimum. Several heuristics allow such local optimizations to scan a wider range of the variable space.

Simulated annealing [91] perturbs the found set of locally optimal variables in the hope of moving it to the attracting basin of another, potentially better optimum. Sampling strategies, such as the systematic ***Latin hypercube sampling*** (LHS, [120]), start the optimization from many different initial points, to allow scanning of possible optima in the entire variable space. For a large number of variables, such approaches require usually infeasibly many optimization runs to reasonably cover the variable space. Other global optimization heuristics borrow principles from genetics, evolution or particle swarms to escape local optima, all based on stochastic elements. In practice, such algorithms usually show poor convergence [155].

Multiple objectives

Biological systems often have to fulfill more than one objective. Microbial metabolism, for example, can be described as trade-off between optimizing energy-yield, growth rate and enzyme investment [179]. In such cases, optimization problems are called ***multi-objective*** and their solutions will lie on ***Pareto fronts***. Each point on the Pareto front describes one optimum solution, for which no single objective can be improved further without another objective becoming worse. In practice, multi-objective functions are often weighted sums of objectives. The weights assigned to each objective denote the importance of their contribution to the overall optimality and hence changing weights can move the Pareto front.

2.3.4 Optimization of biological models

The model calibration problem

For biological ODE models (Equation 2.3), many of the parameters in $\boldsymbol{\theta}$ are usually unknown and have to be optimized such that the model will show a biologically meaningful behavior. In such cases, the objective function g (Equation 2.20) will describe the distance of the model trajectories to some experimentally measured data points $\hat{\mathbf{y}}$.

While the model usually describes the behavior of the internal biological species $\mathbf{x}$, experimental measurements cannot always access each species individually. Often, only sums of species, such as isomeric metabolites or different phosphorylation states of proteins, can be measured. To formally address this problem, the concept of ***observables*** has proven useful. Observables $\mathbf{y}$ are functions of internal variables and specific parameters $\boldsymbol{\theta}^e$

$$\mathbf{y}(t,\boldsymbol{\theta}) = f(t,\mathbf{x}(t,\boldsymbol{\theta}),\boldsymbol{\theta}^e). \tag{2.23}$$

$\boldsymbol{\theta}^e$ are thereby the parameters of the error model σ associated with the observable, f is an arbitrary function such as a sum. The ***error model*** can be used to encode information about measurement noise in the observable. For biological data, usually a log-normally distributed error is assumed. Other common forms of error models are fractions of the observables (*e.g.* the error is 25 % of the observed value) or additive offsets (*e.g.* corresponding to a blank measurement). In this work, the error model was estimated alongside the ODE model. This approach is opposed to directly calculating an estimator such as the standard deviation from the

data, which has been shown to improve the parameter estimates' accuracy [155].
The distance between model observables $\mathbf{y}$ and measured data $\hat{\mathbf{y}}$ can be formalized as the sum of squared differences between them. To also include measurement noise, the ***likelihood*** function $\mathbf{L}$ can be used as objective g:

$$\mathbf{L}(\hat{\mathbf{y}}|\boldsymbol{\theta}) = \prod_{i=1}^{n_t}\prod_{j=1}^{n_y} \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} exp\left(-\frac{1}{2\sigma_{i,j}^2}\left(\hat{y}_{i,j} - y_j(t_i,\boldsymbol{\theta})\right)^2\right), \tag{2.24}$$

with n_y as the number of observables and n_t as the number of measured time-points.
The likelihood describes the conditional probability of the model with the parameter set $\boldsymbol{\theta}$, given the measurements in $\hat{\mathbf{y}}$. The likelihood takes this form only for normally distributed errors. To use this notation $\mathbf{y}$ and $\hat{\mathbf{y}}$ can be transferred to the log space, transforming the log-normally distributed error to a normally distributed one. Accordingly, $\mathbf{y}$ and $\hat{\mathbf{y}}$ will denote the logarithmic entities in the following.
The optimization problem can then be written as

$$\min_{\boldsymbol{\theta}} \; -2\ln(\mathbf{L}(\hat{\mathbf{y}}|\boldsymbol{\theta})) = n_t \cdot n_y \cdot \ln(2\pi) + \sum_{i,j}\left[2\ln(\sigma_{i,j}^2) + \left(\frac{\hat{y}_{i,j} - y_j(t_i,\boldsymbol{\theta})}{\sigma_{i,j}}\right)^2\right], \tag{2.25}$$

with $-2\ln(\mathbf{L}(\hat{\mathbf{y}}|\boldsymbol{\theta}))$ denoting the log-likelihood $\mathbf{LL}$.
What was referred to as the variable space before will now be the parameter space, as all variables to be optimized are parameters of the ODE model. The first term in the sum in Equation 2.25 denotes the individual contribution of the error model σ to the likelihood.

Profile likelihoods and confidence intervals

The likelihood function can also be used to calculate ***confidence intervals*** *CI* for the parameter estimates. Given a confidence level α, these are regions that contain the true parameter value with a probability of α. The confidence region is bounded by the parameter value for which $\mathbf{LL}$ exceeds a threshold Δ_α compared to the lowest likelihood value found. Δ_α can be computed as the α-quantile of a χ^2 distribution, with one degree of freedom (point wise confidence intervals) or as many degrees of freedom as the number of parameters (simultaneous confidence intervals).
For which values CI_i^u and CI_i^l of a parameter θ_i this threshold is reached, can be calculated by ***profile likelihood estimation*** (PLE). The profile likelihood is obtained for each parameter by increasing and decreasing it and refitting all other parameters until $\mathbf{LL}$ exceeds the threshold. In other words, as soon as a perturbation in the optimal parameter value cannot be compensated by adjusting all other parameters, one confidently left the optimum range for that parameter.
The shape of the resulting likelihood profiles can also help to further assess the quality of the estimation for each parameter (Figure 2.2). From the profiles, three general types of ***identifiabilities*** can be characterized: Structural non-identifiabilities (Figure 2.2 A) do not allow for a sensitive estimation of the parameter. They usually hint to explicitly dependent parameters (*i.e.* one parameter can be expressed as a function of another) and can only be resolved by changing the model structure. Practical non-identifiabilities (Figure 2.2 B) mark parameters that have a defined optimum value, which cannot be detected with high enough confidence. They can be resolved by additional measurements. For identifiable parameters the PLE exceeds the threshold in both directions. PLE derived confidence intervals can also be asymmetric, as opposed to classical confidence estimators, representing the nonlinear properties of the system. [56, 97, 154, 155]

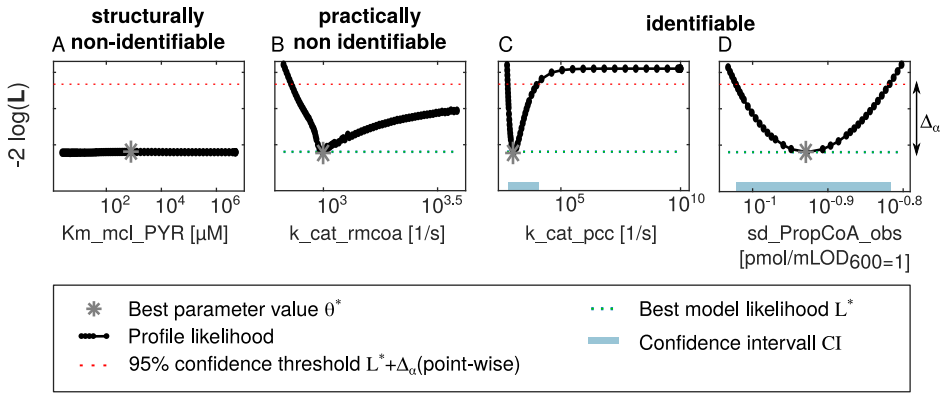


Figure 2.2: Possible scenarios of parameter identifiabilities. **A** Structural non-identifiability with a flat likelihood profile, **B** practical non-identifiability with a defined optimum that is not different enough from the surrounding region. **C, D** Identifiable parameters with asymmetric and symmetric confidence intervals, respectively. The parameter values for which the likelihood profile hits the 95 % confidence threshold mark the upper and lower limit of the confidence region of the parameter (blue interval). Each black dot represents an individual fit. The example is taken from Chapter 7.

2.3.5 Used software

Within this work, the solvers of the Matlab ODE-suite [182], which are based on Runge-Kutta or trapezoidal rules, where used to solve differential equations. For the simulation of the dynamic metabolic models, the Matlab based D2D framework [156] was used, which invokes the CVodes solver [181] in C to speed up the calculations. The solver uses BDF and Adams-Moulton-type algorithms and can additionally supply sensitivity equations. Those can be used to calculate gradients for the optimization algorithms.

For all linear optimizations, the open source solver `glpk` [112] was used with the default simplex algorithm. `glpk` also includes branching sub-algorithms for MILP problems as well as a pre-solving routine to simplify redundant problems.

For constrained nonlinear problems the Matlab internal `fmincon` optimizer was used with an interior point algorithm. Linear as well as nonlinear constraints were implemented. For dynamic model fitting the Matlab optimizer `lsqnonlin` with a trust region algorithm embedded in the D2D tool was used. D2D supplies symbolic sensitivity equations for faster and more accurate calculation of the Jacobian. To address local optima issues, LHS strategies were applied and convergence was monitored with convergence plots as described by Raue *et al.* [155]. Well converging algorithms are thereby characterized by a step-like profile of final objective values, where each plateau can be assigned to a local optimum (for examples see Appendix A.28).

3. Data basis

The experiments described in the following Section were planned and executed within the scope of this work in collaboration with ***Michael Zimmermann****. Metabolomics and proteomics measurements as well as the corresponding sample preparation were carried out by* ***Michael Zimmermann*** *and* ***Olga Schubert****, respectively. Experiments and measurements were conducted at the Molecular Systems Biology Department of the ETH in Zürich. All subsequent data analyses are part of this work.*

This Chapter summarizes the experimental efforts specifically linked to this work. As the experiments themselves are not part of this work, the description remains at a low level of practical detail, referencing to the corresponding literature for in-depth descriptions and reproducibility. The Chapter focuses on a general understanding required for the classification of the strengths and weaknesses of the experimental methods for the modeling approaches central to this work. Additionally, the Chapter includes the non-experimental methods used for processing of the acquired data and preparations for their integration into mathematical models. Thereby, further data from literature and open source databases is used as is described in the following pages.

3.1 Experimental design for modeling

A set of experiments was designed within the scope of this work to decipher the dynamic behavior of the central carbon metabolism in mycobacteria (schematic overview in Figure 3.1). The experiments were intended (i) to cover different carbon sources which require a vastly different flux distribution in the central carbon catabolic network, (ii) to test common experimental carbon sources, such as glucose and glutamate but also to (iii) relate to possible carbon sources that might be available to intracellular bacteria, such as propionate, and (iv) to dynamically challenge the central metabolism to be able to learn about transient flux rerouting and the kinetic properties of the involved enzymes, in a resolution suitable for mathematical modeling approaches.

M. bovis BCG as experimental system

Mycobacterium bovis bacillus Calmette-Guérin (BCG) is the mycobacterial strain used for vaccination against tuberculosis. *M. tuberculosis* and *M. bovis* BCG share 99.5 % of their genome [19], which is why the non-pathogenic *M. bovis* BCG is frequently used as an experimental model system for *M. tuberculosis*. Handling of *M. bovis* BCGdoes not require special security laboratories, whereas experiments with *M. tuberculosis* require biosafety level 3. *M. bovis* BCG, after decades of laboratory life, has lost three genetic regions, mostly including genes for virulence relevant processes. In ^{13}C labeling experiments, minor differences in the

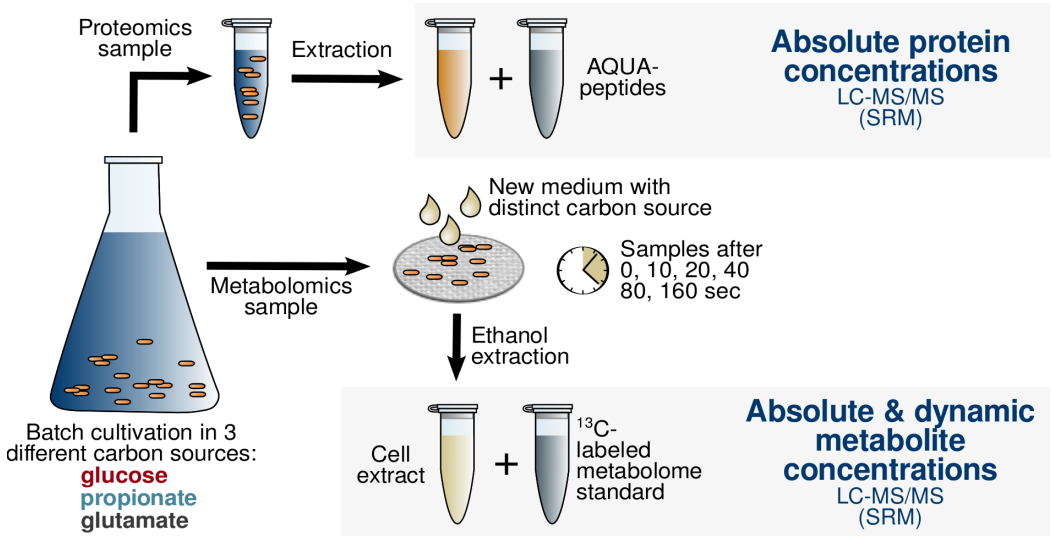


Figure 3.1: Schematic representation of the experimental workflow. Please refer to the text for a detailed description.

metabolic flux distribution between *M. tuberculosis* and *M. bovis* BCG have been found [21], but the overall pattern is conserved, such that insights from *M. bovis* BCG are most likely transferable to *M. tuberculosis*.

Experimental workflow

M. bovis BCG cells were cultured in media containing solely one of the carbon sources glucose, glutamate or propionate, which enter the central carbon metabolism via glycolysis, the Krebs cycle and the methylcitrate and -malonyl pathways, respectively. Samples were taken at optical densities listed in Table 3.1 and transferred to a Büchner funnel with nitrocellulose filter, which is constantly perfused with medium. The cells, which are trapped on the filter, were allowed to acclimate to the new aeration conditions for one minute under constant perfusion with media containing the culture carbon source, before they were perfused with new media containing a different carbon source. The resulting abrupt switch in carbon source availability forces the bacteria to quickly adapt their metabolic fluxes. The intracellular metabolites were extracted at different time-points within three minutes after the shift to obtain time series data. The experiments were carried out as described in Murima *et al.* [129].

Conversions and comparability

The cell number was assumed to scale approximately linear with the optical density and the cell size was taken to be constant over the range of growth conditions tested. Hence, a constant factor of $8 \cdot 10^8 \frac{\text{cells}}{\text{mL}_{\text{OD600=1}}}$ was used to convert between available optical density values and cell numbers. To calculate intracellular concentrations, a cell volume of 0.5 fL was used, approximated from the shape, length [4, 191] and an assumed diameter of $\sim 0.2\ \mu$m (see also Table 4.1). One unit of metabolite or protein in the raw data units $\frac{\text{fmol}}{\text{mL}_{\text{OD600=1}}}$ then translates to 1.25 $\frac{\text{molecules}}{\text{cell}}$ or a 2.5 nM concentration.

3.2 Proteomics

As described in the Introduction (see Equation 1.1), intracellular fluxes depend linearly on the active enzyme concentration. It was therefore intended to gain insight into the cellular protein distribution in the experimental conditions used. Many methods are available for relative measurements of protein concentrations, including the well known Western Blot. In contrast to this qualitative method, mass spectrometry driven quantitative proteomics allow for the absolute quantification of proteins within a cell while yielding a high coverage of the proteome. The absolute concentrations can then be used directly in the rate laws of kinetic models. The genome wide data can be used to gain insight into the general cellular constitution and its enzymatic inventory.

3.2.1 Data acquisition

Samples

For each cultivation condition samples for proteomics measurements were taken (Figure 3.1). It was assumed that the proteome will change with the cultivation carbon source but be approximately constant during the relatively short timespan of the dynamic shift. The analyzed samples are summarized in Table 3.1. The cells were disrupted and subjected to tryptic digestion to obtain peptides that can be screened using mass spectrometry in two different measurement set-ups as described below.

Absolute label free quantification (ALF)

For a proteome wide quantification of the samples, sequential window acquisition of all theoretical spectra (SWATH) mass spectrometry, which is based on absolute label free quantification as described in [177], was employed on whole-cell lysates. The technique applies targeted proteomics (selected reaction monitoring, SRM) using a comprehensive fragment library of all *M. tuberculosis* proteins. The acquired intensity values for each protein, *i.e.* the sum over all associated peptide fragment intensities, were shown to correlate linearly with the actual cellular protein concentration [109]. 27 anchor proteins, spanning the entire protein expression range and being absolutely quantified with peptide standards (see below), were then used to interpolate absolute abundances for all expressed proteins (1527 proteins). This method has the benefit of a genome wide coverage, but relies on the assumption that peptides behave similarly to their neighboring anchor peptides in the mass spectrometer. This assumption might not hold for all peptides, depending for example on the ionization technique used.

Experiment	Replicate	Name	OD_{600}	$\frac{\mu g \text{ protein}}{mL_{OD600=1}}$
glucose, early	A	Glc1	0.36	105.2
	B	Glc2	0.33	116.0
	C	Glc3	0.36	124.3
propionate, early	E	Prop1	0.35	122.9
	F	Prop2	0.34	141.6
	G	Prop3	0.36	126.8
glutamate, late	A2 #6	Glut1 late	0.93	56.2
	B1	Glut2 late	0.94	49.7
	B2	Glut3 late	0.96	53.1
glucose, late	C1 #2	Glc1 late	0.87	41.0
	C1 #4	Glc2 late	0.90	49.2

Table 3.1: Overview of the used samples for the proteomics measurements

Absolute quantification with standard peptides

For a subset of the central metabolic proteins, stable heavy isotope-labeled reference peptides (AQUA peptides) were available which allow for direct, absolute quantification by SRM mass spectrometry as described by Schubert *et al.* [176]. The AQUA peptides are added to the sample in a known concentration and undergo all processing steps in the same way as the cellular peptides. Biases, for example due to a different efficiency of the ionization, can therefore be corrected after the measurement, resulting in a more accurate estimate of the cellular protein concentration.
Overall, a good correlation between ALF and AQUA data could be observed, except on the boundaries of the expression spectrum for very high or very low abundant proteins (Appendix A.1 A). Hence, AQUA data was preferred in the following and only for proteins not contained in the AQUA datasets, ALF values were used. Both methods can reliably detect a 1.6-fold difference in a protein's concentration between two samples, an accuracy high enough to feed the modeling approaches in the following Chapters.

3.2.2 Data processing

Normalization to total protein amount

The protein quantification is dependent on the efficiency of the protein purification from the bacterial samples. The total amount of protein extracted varied significantly between the samples (see Table 3.1, last column), with samples from later growth phases yielding less protein per mL at a given optical density. This could for example be due to structural rearrangements within the cells restricting the protein extraction. The protein concentrations were subsequently normalized to the median of the distribution of the ALF data. In Appendix A.1 B, a summary of the normalization is shown, alongside a comparison between normalizing to the median of AFM or AQUA data, highlighting the influence of the much smaller sample size of the latter.
The normalization neglects the fact that the total protein content (as a fraction of the dry cell mass) could change with the growth phase and rate of the bacteria. However, for *M. bovis* BCG in chemostat culture the protein content does not vary significantly with the growth rate (22.9 % and 21.4 % at doubling times of 69.3 and 23.1 h, respectively [20]). In modeling terms, a change in overall protein content would uniformly scale the measured protein concentrations and accordingly all reaction fluxes, conserving their relation to each other as well as the overall metabolic behavior.

Calculation of active site concentration

For modeling purposes, the best practice is to define the catalytic activity k^{cat} per active site of an enzyme complex (*e.g.* [46]). The complex structure was collected for all enzymes of the reactions considered in the modeling approaches. Information on the different enzyme complexes was collected from primary literature and related databases such as BRENDA [30], KEGG [84] and Tuberculist [101]. The results are summarized in Appendix A.20.
To attain the concentrations of active sites, the lowest concentration of a subunit divided by its stoichiometry in the complex was used. This assumes that all the enzyme measured is active in one reaction, neglecting potential moonlighting activity in other pathways [81]. Therefore, if a protein exhibits significant moonlighting activity, its k^{cat} will be underestimated in the reaction in question, whereas all other parameters should remain valid. For reactions that are catalyzed by two or more isoenzymes, their active concentration was added and they were assumed to have identical kinetics. To resolve differences between isoenzymes, additional data, *e.g.* from knock-out experiments, would be required.

3.3 Metabolomics

Intracellular metabolite concentrations are required to gear metabolic models to biological reality. Their absolute quantification can be achieved with high accuracy by targeted mass spectrometry. The obtained values can be directly included in and compared to the model simulations. Metabolic fluxes can be reconstructed from such data, when direct flux measurements, for example via ^{13}C-labeling experiments, are not available as is the case for this work. To identify the correct order of magnitude of intracellular fluxes, uptake and secretion rates of a batch cultivation experiment were integrated as described below.

3.3.1 Data acquisition

Samples

All shift experiments for which metabolomics data were acquired are summarized in Table 3.2. Metabolites were measured at 0, 10, 20 40, 80 and 160 seconds after the switch to the new medium (see Figure 3.1). Control experiments were carried out analogously, but with media containing the same carbon source before and after the switch.

Absolute quantification by mass spectrometry

Metabolites were measured by mass spectrometry and absolutely quantified by relating them to ^{13}C-labeled standard metabolites with known concentration as described by Büscher *et al.* [26]. Similarly to the proteomics method, a ^{13}C-labeled standard with a known concentration for each desired metabolite is added to the probe after sampling and undergoes all processing steps together with the internal metabolites. All influences of the subsequent extraction and measurement steps will will distort standard and bacterial metabolite pools in the same way, such that the original bacterial concentrations can be recalculated with the help of the measured

Table 3.2: Summary of the metabolomics data.

Cultivation on	Growth phase	Shift medium	# Replicates
glucose	early	acetate	1
		glucose	3
		glucose, acetate	3
		glucose, glutamate	3
		glucose, glycerol	3
		glucose, propionate	3
		glucose, pyruvate	1
		glucose, succinate	1
		glutamate	1
		glycerol	1
		no carbon	1
		propionate	1
		pyruvate	1
		succinate	1
propionate	early	glucose	1
		glutamate	1
		propionate	1
		propionate, glucose	1
glucose	late	glucose	1
		glucose, acetate	1
		glucose, glutamate	1
		glucose, glycerol	1
glutamate	late	glutamate	3
		glutamate, acetate	3
		glutamate, glucose	3
		glutamate, glycerol	3

standard concentration. In total, concentrations of 35 different metabolites participating in the central carbon metabolism could be measured, some of which were pooled concentrations of two or more metabolic species (Appendix A.19). Some metabolites are only available for a subset of the measured samples. On average, a standard deviation of 25 % can be assumed for measurements of that type. [26, 129]

3.3.2 Estimating the range of boundary fluxes

Unfortunately, no direct measurements of the uptake and secretion rates of the above cultures were available. Hence, physiological ranges for uptake and secretion fluxes were estimated from a reference experiment of *M. bovis* BCG grown on glutamate and glycerol (experiment by courtesy of Michael Zimmermann at the ETH Zürich). An exponential growth model was fitted to the bacterial growth curve and corresponding uptake rates for all measured compounds (glycerol, glutamate, citrate, isocitrate, aconitate, succinate) were estimated based on it (Appendix A.2). Based on the estimated rates, the general boundaries for growth, uptake and secretion rates were defined as summarized in Table 4.2.

3.4 Notes on enzyme kinetics

Enzyme kinetics and their parameters have been measured in a vast number of experimental biochemistry studies [196]. In such assays, enzymes are purified from cell extracts, exposed to different concentrations of their substrates and their rate of substrate consumption is measured. From these data, kinetic parameters can be calculated, as in the pioneering work of Michaelis and Menten [122]. Ideally these parameters could be incorporated directly as parameters in ODE models, but in practice they often prove incompatible with the modeled *in vivo* conditions or lack mechanistic resolution.

On the one hand, *in vitro* conditions such as pH, availability of cofactors, salt concentrations or the lack of chaperones in the essay medium, can have a large impact on the enzymatic turnover rate. The assay mixture also cannot mimic possible intracellular regulation. On the other hand, different notations are used to describe enzyme action, some of which are not directly integrable, as they do not normalize for mass or amount of the enzyme tested, such as for example the specific activity in μmol/(min $\cdot$ (mg total protein)). In addition, conversions between different measures are non-trivial and accordingly converted values as well as database entries can be error prone [46].

k^M values are somewhat less affected by these error sources, as the tested concentrations are set externally, such that they can be used as a direct prior for model parameters. A list of k^M values available for *M. tuberculosis* and other related mycobacteria can be found in Appendix A.15.

III Results

4. Explore the data

Besides many other hopes placed on systems biological approaches, the integration of virtually all available types of data on a biological system is one of the major promises of this scientific field. While carefully considering the experimental contexts and accordingly their potential limitations, bringing together all available data can give an unbiased view on the current state of knowledge on the specific system. Usually, contradictions will occur that can either point to experimental inconsistencies or, more excitingly, reveal knowledge gaps that require closing in order to fully understand the underlying biological process. Mathematical models can be used to bridge these gaps with several hypotheses, often identifying a most likely one which is then open for targeted experimental validation.
The following Chapter will deal with the careful assessment of the data situation, integrating the acquired experimental measurements with other data available from literature. The contextualized data will be the basis for the studies presented in the subsequent Chapters and outline the insight that can be gained from the data alone, without any modeling involved.

4.1 Measuring *M. tuberculosis*

M. tuberculosis by the numbers

As introduced before, quantitative biology is constantly improving itself, yielding data covering ever-larger parts of the biological system in question with ever-higher accuracy. While, for example, individual protein concentrations used to be assessed relatively by the density of Western Blot bands, they can now be measured with quantitative mass spectrometry for the

Table 4.1: Summary of important numbers and characteristics of *M. tuberculosis* as used throughout the work. BNIDs are identifiers of the Bionumbers database [123].

Entity	Value	Unit	Reference
cell length	1.5 - 6	μm	(40, 168)
cell diameter	0.2 - 0.5	μm	(40, 44)
thickness of cell envelope	15 - 20	nm	(44)
intracellular volume	0.5	fL	calculated from shape (lower bound)
elemental composition (dry mass)	$C_1H_2N_{0.1}O_{0.7}P_{0.005}S_{0.002}$	-	(20)
mass per mol carbon	26	g/Cmol	calculated from formula
mass of one cell	216.5	fgDW	calculated based on (20)
mass of all cells in 1 mL culture at OD_{600} = 1	343.4	μgDW	measurement M. Zimmermann, ETH Zürich
number of cells in 1 mL culture at OD_{600} = 1	$1.58 \cdot 10^9$	-	calculated
division time on glycerol and glutamate	38	h	BCG reference dataset
total carbon uptake	135	μM C/s	BCG reference dataset
total carbon required to build one cell	8.3	fmol C/cell	calculated
total carbon flux required to build one cell	121.2	μM C/s	calculated
intracellular pH	7.5	-	assumed, BNID 106187,108096,111346
total intracellular ionic strength	0.2	M	assumed, BNID 107522

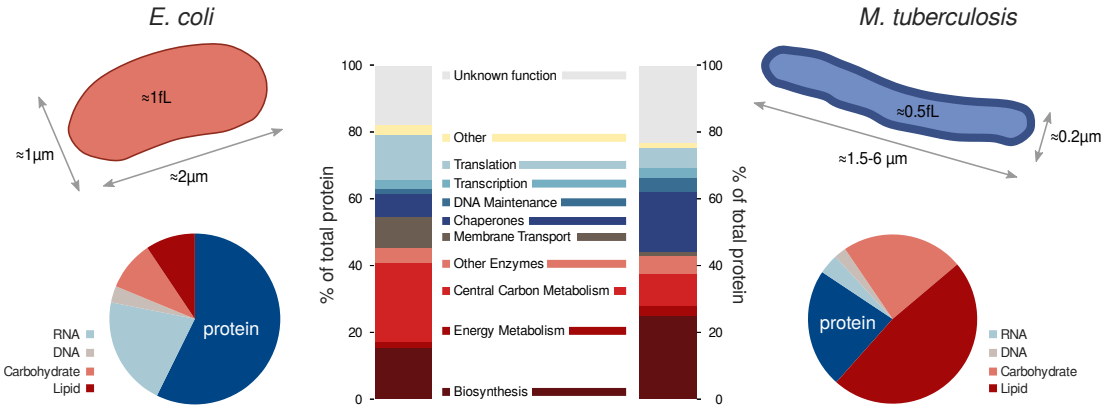


Figure 4.1: *E. coli* and *M. tuberculosis* compared. Upper part shows a schematic drawing of the cell shapes, with the corresponding dimensions. The pie charts below denote the macromolecular composition of the cells. The stacked bars dissect the protein fraction further into functional categories. Percentages refer to mass percent (%(w/w)). The categories are further decomposed in Figure 4.2 A. *E. coli* data is taken from Bionumbers (BNID 101788, 111490).

entire proteome. The calculation of absolute abundances from such data, however, has to be embedded in the overall mass and energy balance of the cell, which is why general features such as cell size and mass have to be considered during the calculations. Based on those numbers, it is also possible to estimate sensible ranges for biological entities. For example, the range of the carbon uptake flux can be estimated from the molecular composition of the cell and its growth rate. For *M. tuberculosis*, the relevant published and derived numbers are collected in Table 4.1.

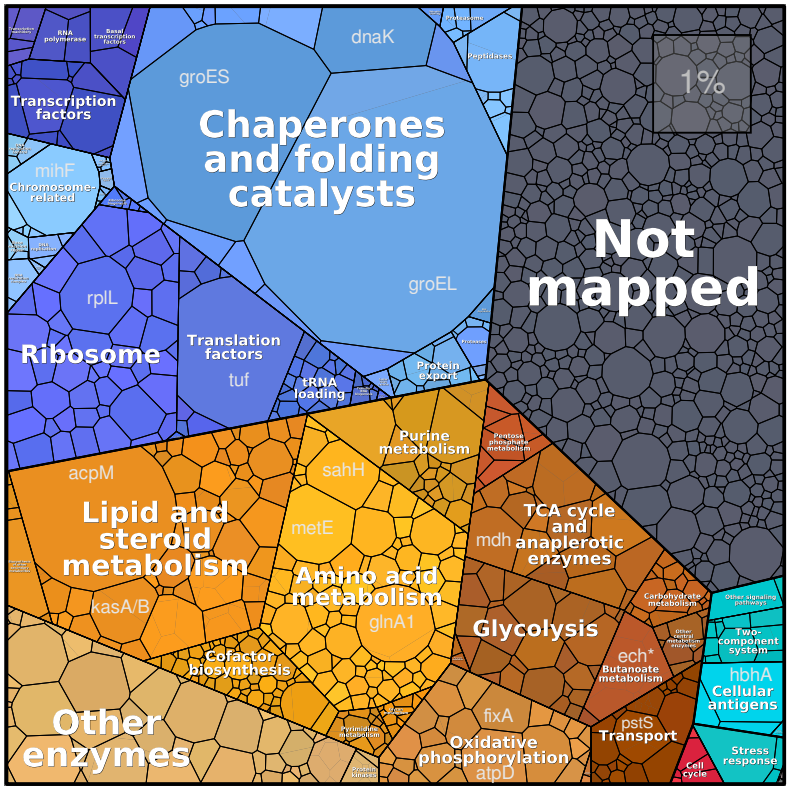
M. tuberculosis compared to other cells

For context, the physiology of *M. tuberculosis* as described by the collected numbers above, will be related to other cells in this Section. Compared to its natural habitat, the human macrophage with a volume of roughly 5000 fL (BNID 103566), *M. tuberculosis* is four orders of magnitude smaller (Figure 1.2). Nevertheless, the bacteria are often densely packed in the phagosomes of infected macrophages, resulting in nutrient and spatial limitations.

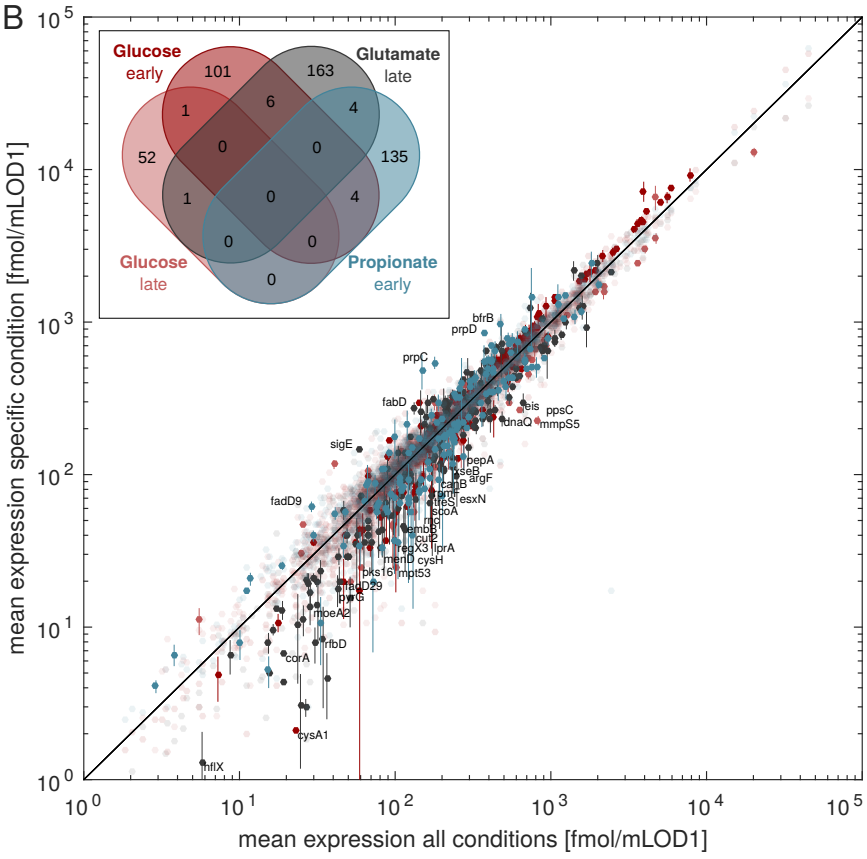
While comparable in size, the surface to volume ratio of *M. tuberculosis* is larger than that of the model bacterium *E. coli* (Figure 4.1). Along with the thick cell envelope, this poses a substantially higher biosynthetic burden for cell wall biosynthesis on *M. tuberculosis*. This is also reflected in the cellular composition of the bacterium, whose dry weight consists to almost 45 % of lipids, required in large number to form the cell envelope [20]. Additionally, complex carbohydrates are incorporated in the envelope. They constitute 20-30 % of the cell dry weight, depending on the growth rate.

Figure 4.2: **A** Proteomap [105] of the *M. bovis* BCG protein composition in propionate media. The area representing 1 % of the total protein number is shown in the upper right corner. Functional classes are grouped and colored together, single tiles represent the amount of a protein. Some selected proteins of each class are labeled with their common name. **B** Ranges of protein concentrations and their variability between the four conditions compared to the mean expression in all conditions. Each dot represents one protein, differentially expressed proteins are shown in non-transparent color (dark red - glucose, early growth phase, blue - propionate, early growth phase, light red - glucose, late growth phase, gray - glutamate, late growth phase). Inset: Venn diagram of differentially expressed genes ($p<0.01$, [1]). Please see Methods Section 2.1.3 for criteria of differential expression.

A


dnaK
groES
Chaperones and folding catalysts
1%
Transcription factors
mihF
Chromosome-related
Not mapped
groEL
rplL
Ribosome
Translation factors
tuf
tRNA loading
Protein export
Purine metabolism
acpM
sahH
Lipid and steroid metabolism
metE
TCA cycle and anaplerotic enzymes
mdh
Amino acid metabolism
kasA/B
glnA1
Glycolysis
Carbohydrate metabolism
Two-component system
ech*
Butanoate metabolism
hbhA
Cellular antigens
Cofactor biosynthesis
Other enzymes
fixA
pstS
Transport
Oxidative phosphorylation
atpD
Stress response
Cell cycle

B


Glucose early
Glutamate late
Glucose late
Propionate early
101
163
52
135
6
4
1
0
0
1
0
4
0
0
0
mean expression specific condition [fmol/mLOD1]
mean expression all conditions [fmol/mLOD1]
bfrB
prpD
prpC
fabD
ppsC
mmpS5
sigE
pepA
fadD9
esxN
iprA
cysH
pks16
mpt53
mbeA2
rfbD
corA
cysA1
hflX

For bacterial cell division, especially the number of ribosomes plays an important role. From the proteomics data (see Section 3.2), the fraction of ribosomal proteins in the total protein amount can be calculated as roughly 3 % (Figure 4.2 A). The fast growing *E. coli*'s ribosomal fraction is higher, reaching up to 20 % at high growth rates (BIND 102345). The ribosomal fraction, along with the cellular translation rate defines the translational capacity, which correlates well with the division time of bacterial cells ([180], Appendix A.3). Accordingly, due to the low percentage of resources allocated to the replicative machinery of the cell, *M. tuberculosis* is limited to the slow growth rate observed experimentally. The majority of resources seems to be rerouted to build up defensive biomass in the cell envelope.

4.2 Analysis of the experimental data

In the following Sections, the protein and metabolite data obtained from the shift experiment are analyzed. Metabolic setup and protein inventory are compared between the experimental conditions. A first scan of the thermodynamic landscape based on the reversibility index is conducted. While not employing any mathematical modeling, conclusions about proteome and metabolome homeostasis as well as on potential thermodynamic bottleneck reactions can be drawn.

4.2.1 Proteomics

Functional proteome fractions

First, the overall distribution of the mycobacterial proteome is analyzed. For that, the distribution of proteins, clustered according to their molecular function is summarized in Figure 4.2 A, exemplarily for the propionate culture condition. The most abundant proteins are the chaperones groEL and groES. Together with other proteins of that class, they form a large group of non-growth associated proteins. In contrast, ribosomal proteins make up only about 3 %(w/w) of the total proteome, resulting in the low growth rate as described before. In a similar way, catabolic enzymes, generating energy and biomass precursors (brown tiles in the Figure, 12.3 %(w/w)) are outnumbered by large amounts of biosynthetic proteins, especially for lipid and amino acid synthesis (beige/yellow tiles, 25.1 %(w/w)), which require precursor supply from the former group. Only a small fraction of the proteome is dedicated to regulatory functions (cell cycle, red; transcription factors, dark blue; signaling, turquoise). A rather large fraction of the overall protein has unknown function (gray tiles, 23.1 %(w/w)).
From this functional proteome distribution, mycobacterial protein resource allocation can be characterized as being governed by protective efforts such as protein maintenance via chaperones or the biosynthesis of protective lipids. Catabolic enzymes, energy generating reactions and ribosomal proteins, which fuel and facilitate cellular growth, receive little resources, correlating with the bacterias' slow growth rate.

Differential expression

The overall proteome composition is well conserved between the four analyzed conditions, with less than 200 proteins expressed differentially in each condition (Figure 4.2 B). The sets of differentially expressed proteins show very little overlap between the conditions (Figure 4.2 B, inset). However, it has to be stated here that the statistical power of the differential expression test is low, with only 11 samples per protein from the proteomics measurements (see Table 3.1). Unsurprisingly, GO-enrichment tests identified that for all conditions metabolic proteins are over-represented in the differentially expressed set (Appendix A.4), including processes of nucleotide, fatty acid and amino acid biosynthesis. When the amino acid glutamate is supplied

in the medium, the processes surrounding that metabolite, such as arginine and glutamine biosynthesis, appear to be regulated by differential expression. In summary, the bacteria seem to maintain their general proteome composition in changing nutrient conditions, at least for the carbon sources tested here, while only a small subset of metabolic genes change their abundance.

The central metabolic proteins

The central metabolic enzymes are of special importance for the following Chapters, which is why their concentration pattern in the different nutrient conditions is examined here in detail (Figure 4.3 A). As described in the Methods Section, proteomics samples were normalized to the median of the ALF quantification. However, even after this normalization, overall protein expression in the central carbon metabolism differs between the conditions, potentially masking relative changes of the proteins to each other. More specifically, on propionate media, enzymes of the central carbon metabolism seem to be generally higher expressed than in other conditions (red color in Figure 4.3 A). It is unclear if this is a real biological effect, a side effect of the normalization or due to measurement error. Nevertheless, some protein concentration changes are evident even given this limitation. Specifically, Pep-carboxykinase (pckA) is up-regulated in the two gluconeogenetic conditions, consistent with its preferentially gluconeogenetic function described *in vitro* [110]. The proteins of the pentose phosphate pathway appear to be higher expressed in early growth phases. This is especially evident from the comparison between the two glucose conditions, arguing for a carbon source independent regulation of the pathway. Under propionate conditions, the initial enzymes of the Krebs cycle (citrate synthase, aconitase) do not follow the overall up-regulation trend, which could hint to a specific regulation of the amount of flux entering the cycle in this condition. The enzymes of the methylcitrate pathway are highly up-regulated under propionate conditions, consistent with their catabolic function. The parallel methylmalonyl pathway does not show an equally strong increase between glucose and propionate conditions (and was not detectable in late growth phase samples).
In summary, some of the observed proteome changes in the central carbon metabolism can be directly related to the change in nutritional condition, such as the up-regulation of the methylcitrate pathway enzymes on propionate or the increased concentration of pckA during gluconeogenesis. Reasons for other changes remain elusive, such as the lack of up-regulation of the methylmalonyl pathway on propionate, but can be tackled by modeling efforts in the following.

4.2.2 Metabolomics

Condition dependent metabolite concentration changes

In comparison to the protein measurements, the metabolomics data show clearer condition dependent concentration changes (Figure 4.3 B). When classified by their steady state, measurements of the late growth phase samples from the cultures on glutamate and glucose show similar metabolite landscapes. Concentrations in those two conditions are very distinct from the early growth phase samples from propionate and glucose cultures, which in turn are clearly distinguishable from each other (based on principle component analysis, Appendix A.5). Besides the nutrient availability, the growth phase therefore seems to have major influence on the intracellular metabolic landscape. Several metabolites show distinct concentration levels for a specific condition or growth phase (Figure 4.3 B and in detail in Appendix A.6). The pentose phosphate pathway intermediates ribose- and ribulose-5-phosphate (R5P, RL5P) are present in concentrations vastly different between early and late growth phases, consistent with the

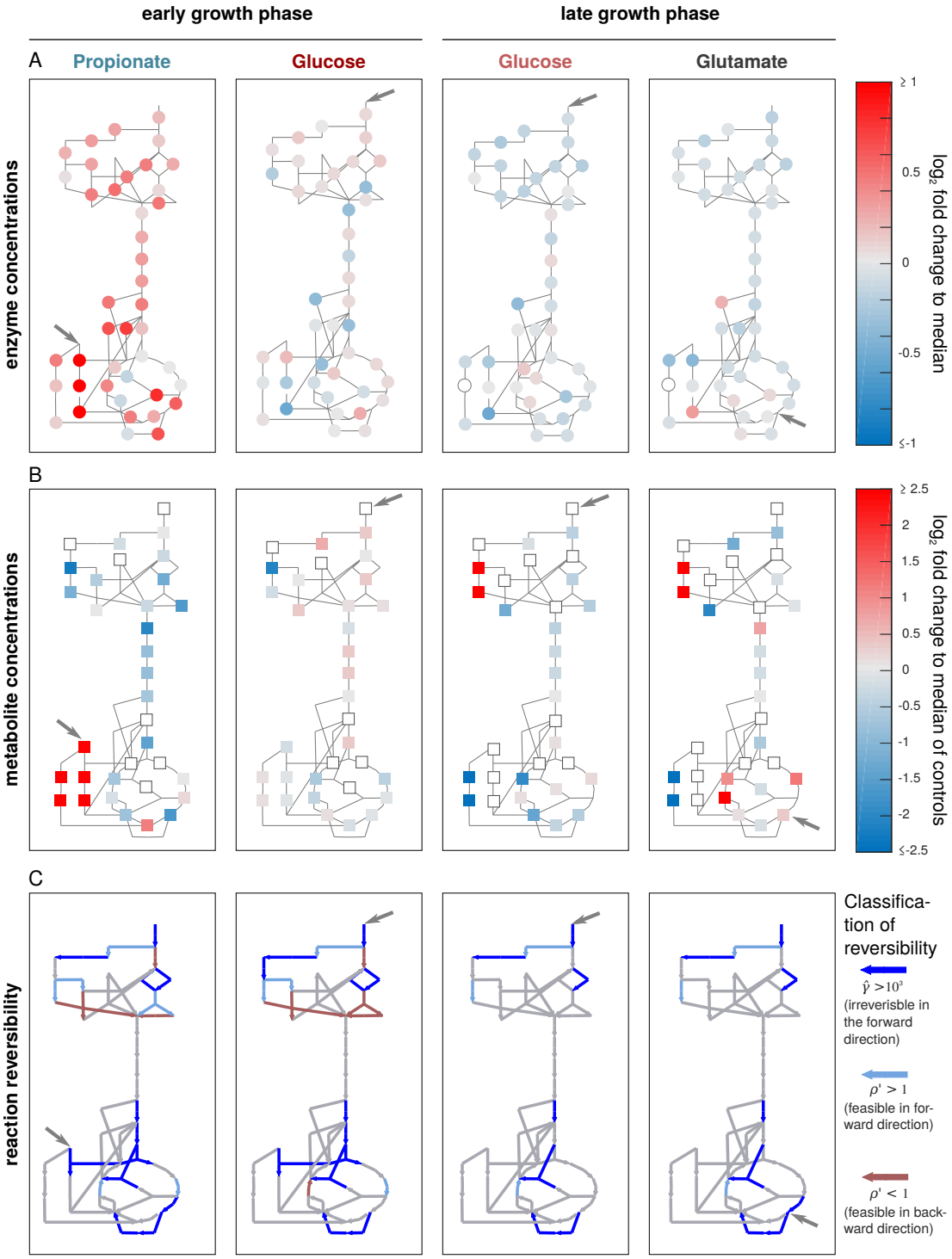


Figure 4.3: **A** Changes in protein concentrations relative to the median of all conditions. Please, note that most of the fold changes have an absolute value smaller than one. Empty circles denote not detected proteins **B** Changes in metabolite concentrations relative to the median of all conditions. Empty squares denote missing values. **C** Reaction reversibilities based the metabolite concentrations defined by the reversibility index $\hat{\gamma}$ (dark blue) or the inverse disequilibrium ratio ρ' (light blue and red). For the calculation of the disequilibrium ratio, the mean steady state concentrations were used. For an annotated map of the depicted reaction network, please refer to Figure 6.1.

observed differential expression of the pathway's enzymes as described in the proteomics Section above. The Krebs cycle shows increased concentrations under growth on glutamate, along with increased levels of the first metabolite of lower glycolysis (bisphosphoglycerate, BPG). Under early phase growth on glucose, ATP and GTP levels are increased (Appendix A.6), along with the concentrations of many glycolytic intermediates, potentially indicating a better overall energetic situation of the cells. In media containing propionate, levels of the intermediates of both available catabolic pathways for propionate (methylcitrate and methylmalonyl pathways) are increased, while the concentrations of other measured metabolites are generally decreased. As changes in metabolite concentrations do not necessarily translate one-to-one into changes in the cellular flux distribution, modeling approaches are used to describe their influence in the following Chapters, explaining the observations described here.
Aside from the steady state behavior, it is also worthwhile to identify metabolites that specifically change their concentration in response to the fast nutrient switches (Appendices A.7 and A.8). Those can hint to transient rerouting strategies of the cells to transit from one steady state to another. Overall, only few metabolites show a strong dynamic change and, not surprisingly, such metabolites will mostly occur in proximity of the perturbation in the metabolic network. In such a way, the intermediates of the methylmalonyl and methylcitrate pathways quickly increase after the addition of propionate to the media. Upon acetate addition, a slow increase of the first metabolites of the Krebs cycle can be observed (aconitate, isocitrate). Dihydroxyacetone phosphate (DHAP) accumulates upon the addition of glycerol.
However, also longer range effects can in distant network parts became evident, such as the decrease of acetyl-CoA upon the addition of propionate and, *vice versa*, the decrease of propionyl and methylmalonyl-CoA when acetate is added to the medium. When the cells are shifted from glucose to acetate only, many intermediates of the pentose phosphate pathway and upper glycolysis are depleted, indicating another network effect. The dynamic data also allow for better assessment of measurement quality. RL5P, as an example, is clearly up-regulated in the late growth phase control conditions as described above, but shows a noisy behavior in the dynamic dataset. This becomes evident when comparing the initial time-point of the dynamic experiment to the control conditions (Appendix A.8, upper left, t = 0s): RL5P shows strong variation over almost all experiments at this time already, potentially arguing for a large influence of secondary effects of the experiment, such as the changed aeration condition on the filter. All changes observed at this time point cannot be due to specific flux-rerouting, as the cells were never exposed to medium with a new carbon source.

General bounds for entities in modeling approaches

According to the range of measured metabolite concentrations and based on the uptake rates from the BCG reference experiment as described in the Methods, general bounds for the entities used in the later modeling process were defined as summarized in Table 4.2.

Entity	Description	Lower bound	Upper bound
$\mathbf{x}$	metabolite concentrations	1 μM	50 mM
k^M	Michaelis Menten constant	1 μM	50 mM
k^{cat}	catalytic activity	10^{-6} 1/s	10^6 1/s
$\mathbf{r}_{up}, \mathbf{r}_{sec}$	uptake / secretion rate	10 μM/s	50 μM/s
$\mathbf{r}_g$	growth rate*	0.017 1/h	0.02 1/h

Table 4.2: Default bounds of entities in the optimization problems based on available data. *Note that the unit is changed to $1/s$ in later analyses for numerical convenience (explained in detail in Chapter 5.1).

Reversibility index: Constraining reaction directionalities

As a first glimpse on the possible flux distribution within the central metabolic reactions, their reversibility was classified individually based on the measured metabolite concentrations with the reversibility index $\hat{\gamma}$ ([133], for detailed description see Methods around Equation 2.14) Whenever measured metabolite concentrations were available for the educts of a reaction, they were included in the calculation of $\hat{\gamma}$. To include uncertainties in the measured metabolite concentrations, the mean measured concentration plus the measurements standard deviation was used if the metabolite occurs as a substrate (accordingly, minus the standard deviation for products). Due to the adjusted default concentration bounds (Table 4.2), a reaction was defined as irreversible here if $|\ln(\hat{\gamma})| > \ln(x_{ub} - x_{lb}) \approx 10.8$. For absolute values of $\ln(\hat{\gamma})$ larger than that, the metabolite concentrations would have to exceed the default concentration bounds to revert the reaction. The default concentration C was defined as the geometric mean of the default metabolite range ($C \approx 220\mu$M). For reactions with all reactant concentrations measured, the inverse of the disequilibrium ratio as $\rho' = 1/\rho = k_{eq}/Q$ (see Equation 2.13) was calculated as a measure for the distance to the equilibrium based on the mean of the metabolite concentrations. As for the reversibility index, ρ' is larger than one if the reaction runs in forward direction. The further it is from one, the more irreversible it will be. The calculated values for $\hat{\gamma}$ and ρ' are mapped to the network in Figure 4.3 C.

Based on this calculations, some reactions can already be associated with a direction. The pyruvate kinase reaction (*pyk*) is irreversibly glycolytic in all conditions. Hence, the only feasible way to feed gluconeogenesis is via PEP-carboxykinase (*pck*), which is in accordance with the found preferential working direction of this enzyme [110] and its increased expression in the proteomics data in gluconeogenetic conditions (see Section 4.2.1). The initial reactions of the pentose phosphate pathway are irreversibly glycolytic. For the early growth conditions, upper glycolysis seems to be thermodynamically restricted.

From the overall distribution of $\hat{\gamma}$ values it becomes evident that the method is heavily dependent on the number of metabolite concentrations known for the specific reaction (Figure 4.4). A reaction is more likely to be irreversible if all its educts concentrations are known. Reversibilities could hence often be due to the lack of constraints from missing metabolite measurements, rather than a real close-to-equilibrium reaction. A bold conclusion of this would be that the majority of reactions can only operate in one direction given a specific metabolic condition. Accordingly, the late growth phase landscapes are less restricted as they are based on a lower coverage metabolite data set.

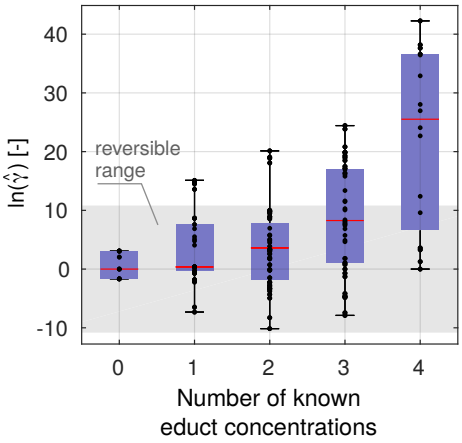


Figure 4.4: Dependency of the reversibility index $\hat{\gamma}$ on the number of known educt concentrations. Reactions, for which all educt concentrations were measured, were left out. Red lines indicate the median, the blue boxes extend between the 25 and 75 % quantiles, whiskers extend between the highest and lowest values, and black dots are all $\hat{\gamma}$ values. The reversible range for $\hat{\gamma}$ values as described in the text is shown in gray.

4.2.3 Discussion

Proteomics

From the data, the proteome of *M. bovis* BCG can be characterized as relatively constant. The cells react specifically by adjusting a small set of protein concentrations in response to perturbations in the carbon source. Metabolic flexibility seems to be limited to a small subset of the metabolic network, at least in the specific experimental setup tested here.
In the face of the functional protein distribution, the slow growth of the bacteria can be explained by the low fraction energy generating and ribosomal proteins. The majority of the resources available for proteins is invested in defensive precautions, such as the production of complex lipids or stabilization of proteins, which do not contribute directly to the replication of the cell but rather to the protection of the cell against the host's defensive actions.

Metabolomics

Steady state metabolite concentrations change dramatically between the tested conditions. Some changes are specific to the growth phase rather than the nutrient condition, such as the concentration increase in RL5P and R5P. The nucleotide triphosphates ATP and GTP are strongly increased in the early glucose condition, potentially arguing for a favorable energetic situation. BPG acts as a marker for gluconeogenetic flux on glutamate, but not on propionate where the same flux reversal is required. In the dynamic data, only few directed metabolite concentration changes could be identified, especially given the measurement noise. The concentrations of most of the metabolites seem to be homeostatic on short time scales. The interconnection of the CoA compounds could argue for a limitation in the cofactor, as soon as larger amounts of acetate or propionate have to be catabolized.

What the data tells

From this purely data based analysis, a lot can be learned about the mycobacterial metabolism, highlighting the strengths of omics approaches. The slow growth rate can be explained by the defensive resource allocation in the proteome. The metabolomics data reveal energetic properties of the different nutritional conditions and allow for the preliminary assessment of thermodynamic features. Metabolic perturbations reveal homeostatic and nutrient- or growth phase-sensitive metabolites.
Modeling efforts can complement these findings by contributing a more detailed analysis of the underlying mechanisms. In the following Chapters questions that remained open in the data analyses will be specifically addressed. The reaction thermodynamics are further assessed and embedded in the network context, resolving inconsistencies on the way (Chapter 5.2). The metabolome and proteome changes are linked mechanistically via a minimum-regulation approach (Chapter 6.2) and on the pathway level, the distinct expression patterns of the two propionate catabolic pathways will be explained (Chapter 7).

5. Linking genome and small scale

How is the central carbon metabolism embedded in the cellular metabolic network?
Which of its reactions are thermodynamically limiting and are they limiting in all conditions?

Systems biology employs two major strategies to understand mechanistic processes (see Figure 1.3). ***Top-down*** approaches measure cellular entities on the whole-cell level: A whole family of omics techniques captures the overall cellular behavior and responses to perturbations. High-level data analysis is then employed to draw conclusions on the underlying regulatory processes. Especially in metabolic research, also modeling efforts have reached the genome size, with models covering the entirety of all known metabolic pathways. On the other side of the complexity spectrum, ***bottom-up*** approaches aim to understand single processes on a detailed level: Classical biochemistry and molecular biology join forces with mathematical modeling to mechanistically describe a single metabolic pathway or a genetic regulation circuit. The mechanistic detail requires a sound data base and comes at a high computational and experimental cost.
This Chapter will address the interface of these two approaches and how knowledge can be transferred from one to the other. Specifically, it will tackle the open issue of how small dynamic metabolic models can benefit from the knowledge collected in genome-scale approaches. A technique is presented that allows embedding of small mechanistic models in the flux context given by a genome-scale model in a specific environmental condition, thereby meaningfully defining their boundary fluxes (Section 5.1). With these calculated leak fluxes, condition dependent, optimum thermodynamic landscapes are then calculated for the central metabolic network in *M. bovis* BCG (Section 5.2), improving the reversibility index approach described in the previous Chapter.

5.1 Biomass backtracking

This Section is based on the publication:
Tummler, K., Kühn, C., & Klipp, E. (2015).
Dynamic metabolic models in context: biomass backtracking.
Integrative Biology, 7(8), 940-951.

5.1.1 Context for small metabolic models

Metabolism has been an early target of mathematical modeling efforts, due to its chemical traceability. Existing models mirror the two modeling types described above: Small, bottom-up models in the form of ordinary differential equations have proven to be powerful tools to

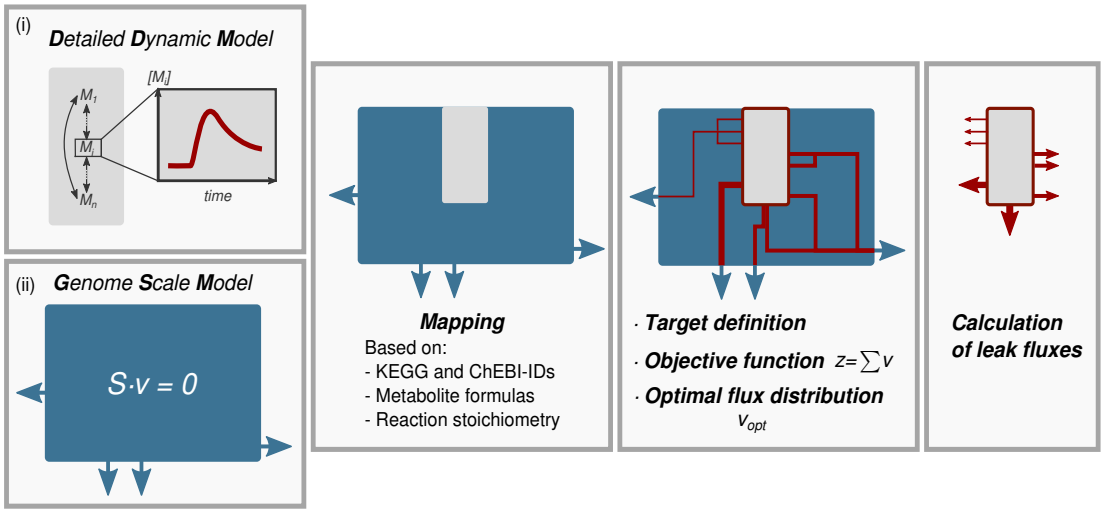


Figure 5.1: Schematic overview of the backtracking approach (Figure from [195]).

understand the intrinsic dynamics and complex functionalities of cellular metabolism [94, 201]. Yet, due to limitations in computational power as well as in the availability of high quality data and calibration algorithms, these algorithms are usually restricted to a small region of the entire metabolic network or to a simplified description. Additionally, single pathways can show drastically different behavior in changing cellular contexts, which might not be included in the focused models. On the other hand, top-down, genome-scale metabolic models can predict genome wide flux distributions under various external conditions integrating a variety of different experimental data [23], but lack mechanistic detail as well as temporal resolution.
A number of methods exist that intend to improve the predictive power of large scale models by linking the two modeling types. Examples reach from the integration of a framing dynamic model describing growth and nutrient uptake rates (dFBA [111]), inclusion of kinetics of single reactions (k-OptForce [34]) to the full "ODE-fication" of the flux carrying sub-network [186] in a specific condition.
The inverse approach of improving dynamic models with genome-scale information is presented here: Biomass backtracking, a method to exploit the extensive knowledge collected in genome-scale reconstructions to contextualize dynamic pathway models. A calculated genome-wide flux distribution is thereby used to establish a stoichiometrically exact, condition dependent biomass reaction for the dynamically modeled network part.
Defining reactions connecting dynamic metabolic models to cellular context, often in the form of biomass reactions, is a complex task that is usually carried out by hand based on one's own experimental data or crude estimates. In this process it is easy to overlook systemic effects such as side products of distinct synthesis pathways that re-enter the dynamically modeled network part (*e.g.* pentose phosphate pathway) or the effect of additional nutrients entering at distinct areas of the metabolic network. However, drains into biomass or cellular products and their exact stoichiometry can have a dramatic influence on the kinetics of the modeled reactions. If, for example, a dynamically modeled metabolite x is consumed in large quantities for the production of biomass, the reactions consuming it in the dynamic model will have to carry much less flux to maintain the steady state concentration of x than the ones producing it. This strongly influences the kinetic parameters of the associated enzymes as well as all downstream fluxes.

Backtracking in a nutshell

The start of backtracking is (i) a detailed dynamic model (DDM) covering part of metabolism but lacking a detailed biomass reaction and (ii) a genome-scale metabolic model (GSM) of the same organism (Figure 5.1). The backtracking workflow embeds the DDM in the broader context of the GSM and provides a stoichiometrically consistent reaction consuming species of the dynamic model to generate biomass (or any other target metabolite) as defined in the GSM. For this, the overlap between the DDM and the GSM has to be defined (mapping). The mapping of metabolites is based on unique database identifiers as well as on metabolite sum formulas. The reactions can be assigned based on identifiers and the reaction stoichiometry linking the mapped metabolites. The mapping will result in a set of reactions and metabolites that is part of both models. Subsequently, a target metabolite from the GSM is selected, for which the required production fluxes from the dynamic model will be quantified. In most cases this will be the biomass, but any other metabolite in the GSM can be used as well, for example if the entity of interest is a specific secreted metabolite. The flux through the GSM reaction consuming the target metabolite is then set to a fixed value, to allow for normalization of the computed leak fluxes. In the next step, a parsimonious flux balance analysis (pFBA, [73]) is run to calculate flux distributions in the GSM. At this point, also further information, such as thermodynamic constraints, experimentally measured uptake rates, fluxes or enzyme expression data, can be included to constrain the solution space of the GSM. From the obtained flux distribution, the leak fluxes of each mapped metabolite of the dynamic model are calculated as the sum of all fluxes affecting that metabolite (i.e. have non-zero entries in the respective row of the stoichiometric matrix), excluding those that have been mapped to be contained in the DDM.

5.1.2 Mathematical formulation

Implementation

The backtracking workflow was implemented in Matlab, employing the SBML toolbox [90] for model import from and export to the commonly used Systems Biology Markup Language (SBML, [78]). The Cobra Toolbox [17] was used to handle models and calculate optimum FBA solutions by means of the open source linear programming solver `glpk` [112].

Preparations: Mapping and target definition

The DDM is mapped to the GSM primarily via annotations of database identifiers (ChEBI [69], KEGG [84]) provided in both SBML files. If annotations are not provided, metabolites can be mapped via their sum formulas, which are automatically looked up in the above databases. The DDM reactions are mapped via KEGG identifiers or similar stoichiometries of mapped metabolites in both models. The mapping is semi-automatized, with a list of likely matching entities provided for each metabolite and reaction, from which the user can select.
One reaction or metabolite has to be defined as the target of the backtracking. The calculated leak fluxes will be normalized to the flux through this reaction, resulting in relative leak fluxes in the unit mol/(mol target). The target will usually be the biomass, but can also be specific (secreted) products of interest or submodules producing individual biomass constituents, as required for modular modeling.

Leak calculation

For the calculation of the leak fluxes a feasible flux distribution for the GSM is calculated. Here, the minimum sum of all fluxes, coinciding with a low overall enzyme load (pFBA, [73]) was used as an objective. pFBA has been shown to correlate well with experimental data [102, 178],

but other more sophisticated or data driven objectives and constraints can be used.
The resulting flux vector $\bar{\mathbf{v}}$, fulfilling the steady state assumption $\mathbf{N} \cdot \bar{\mathbf{v}} = 0$, contains entries v_m for the fluxes of all reactions that have been mapped between the GSM and the DDM, and v_{GSM} for all fluxes only present in the GSM. The leak fluxes $v_{leak,i}$ are calculated as the sum of fluxes v_{GSM} of all reactions that consume or produce a metabolite x_i of the DDM and are themselves not part of it, multiplied by their respective stoichiometric coefficient $\vartheta_{i,j}$:

$$v_{leak,i} = \sum_{j=1}^{n_r} \vartheta_{i,j} \cdot v_{GSM,j}, \tag{5.1}$$

with n_r as the number of reactions in the GSM.

Handling of lumped reactions and metabolites

For reactions and metabolites with one-to-one mapping the leak fluxes can be easily calculated as a simple weighted sum. However, simplifications in the model topology - often by lumping several reactions or metabolites into a common entity - have to be treated with care.
Leak fluxes for ***lumped metabolites*** are calculated as the sum over all leak fluxes of all mapped species of the metabolite pool. Reactions inter-converting between the pool species will cancel out in the summation, but their contribution to the turnover of cofactors will still be represented in the optimal flux distribution. ***Lumped reactions***, for example one single reaction representing a linear pathway, have to be treated more carefully. As they also bridge several intermediate metabolites, the leaks of those will not be represented in the DDM. However, those leaks could be essential sources for biomass precursors and neglecting them could introduce large errors in the flux distribution of the DDM. Therefore, a summed stoichiometry of the lumped reaction from the optimal flux distribution with amended stoichiometric coefficients of substrates and products is calculated. For each lumped reaction, a ***waste term*** is generated including all intermediate leaks from the bridged metabolites.

5.1.3 Results: What backtracking can do for you

Condition and species specific leak fluxes

In the publication [195], several test cases were used to point out the complexity and variability of biomass leak stoichiometry. The impact of altered environmental conditions on the leak fluxes from yeast glycolysis was characterized, based on an experimental data set of yeast grown under different aeration conditions [82]. It was also shown that biomass leaks heavily depend on the organism in question, caused by alternative metabolic wiring or a different biomass composition. Exemplarily, glycolytic leaks for 4 different species with available high quality GSMs (*S. cerevisiae*, *E. coli*, *M. tuberculosis* and *B. subtilis*) showed vastly different distributions.
Backtracking also supplies values for cellular uptake or secretion rates for all required media compounds. This might be helpful if, for example, the sensitivity of a distinct metabolic pathway to the availability of external compounds such as oxygen is to be modeled. Additionally, it allows for a consistency check of the calculated leaks and the GSM flux distribution, *e.g.* by balancing carbon atoms in biomass and uptake fluxes. Additionally, backtracking allows to dissect cofactor turnover, as was shown for the case of ATP in yeast, split in glycolytic and respiratory ATP production in the mitochondrial compartment. These findings are useful if dummy ATP consuming or respiratory chain reactions shall be included in a model, as they allow for an accurate calculation of either stoichiometry.

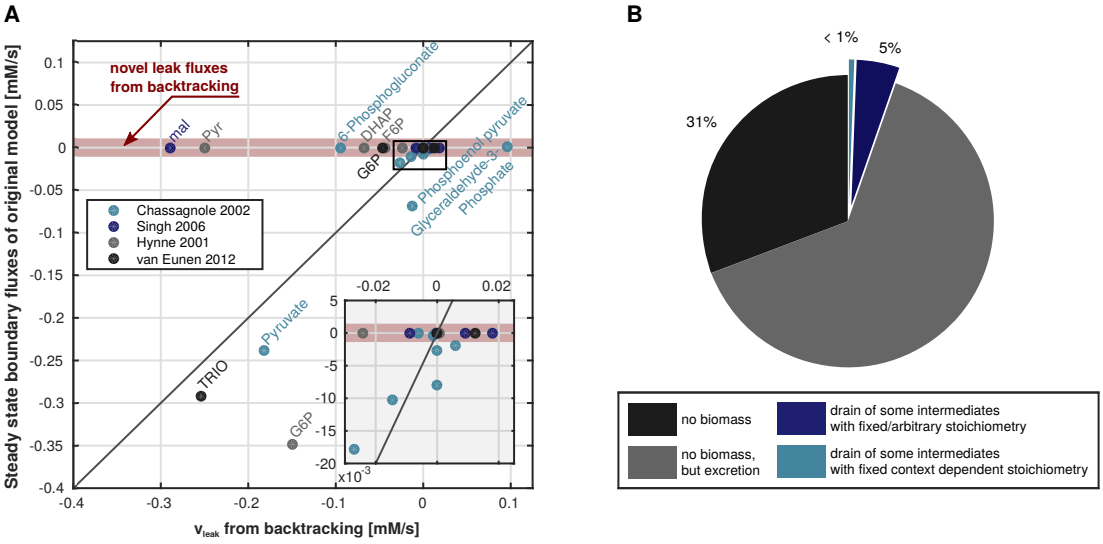


Figure 5.2: Comparison of leak fluxes of existing dynamic models with the results of backtracking. **A** Leak fluxes were calculated for two dynamic models of *E. coli* from Chassagnole *et al.* [31] (BIOMD0000000051) and Singh *et al.* [184] (BIOMD0000000221) based on the iJO1366 GSM [137], and the dynamic models of *S. cerevisiae* from Hynne *et al.* [79] (BIOMD0000000061) and van Eunen *et al.* [200] (MODEL1403250001) based on the Yeast 7 GSM [7] and compared to the steady state fluxes that leave the respective models as simulated in COPASI [75]. Fluxes in the red area were predicted by backtracking but absent in the dynamic models. **B** Summary of all dynamic metabolic models in the Biomodels database with regard to their definition of boundary fluxes. Figure taken from [195], for detailed information on the shown models please refer to the publication.

Most published metabolic models neglect leak fluxes

Despite the importance of contextualization, many published models pay little attention to their boundary fluxes. Of the 169 dynamic ODE models for metabolic pathways in the Biomodels database [32] none has used information on system wide flux distributions, and only 9 take consumption fluxes for biomass production from intermediate species into account (Figure 5.2 B). However, GSMs of over 100 species are available to obtain such information [23].

To exemplify the extent of the leaking fluxes that were neglected in dynamic models, backtracking was applied to several high-quality dynamic metabolic models. Two models of *E. coli* and two of *S. cerevisiae* were embedded in their respective GSMs and leak fluxes were calculated under conditions as defined for the dynamic models. The results were compared to the boundary fluxes present in the original models in steady state (Figure 5.2 A).

For all models, backtracking predicts leak fluxes that were neglected in the DDMs. The closest match to the backtracking leaks was found for the model of Chassagnole *et al.* [31], whose boundary fluxes are based on detailed experimental data. However, the vast majority of currently available metabolic DDMs have less detailed boundary fluxes and could potentially be improved by more accurate leak fluxes from backtracking (Figure 5.2 B).

Model simplification, extension or splitting

Backtracking can also be used to define consistent boundary fluxes for changing model structures in the iterative process of model development or for integrative modeling, where sub-models need to be connected via meaningful interfaces. In the backtracking publication, all potentially occurring scenarios – model splitting, extension or simplification – were exemplified using yeast glycolysis (Figure 5.3 A).

Splitting models into smaller submodules eases computational simulation effort as well as the complexity of parameter estimation, which is why it can be used in large modular modeling projects up to whole-cell modeling. The consistent and automatic recalculation of boundary fluxes between the modules, as exemplarily shown in Figure 5.3 B, is thereby crucial for the interplay of the modules and can be achieved by backtracking with manageable effort.
Simplifying models becomes necessary, wherever the data coverage is not high enough to allow the estimation of kinetic parameters with sufficiently high confidence. The reduction of DDMs can be based on time scale separation (lumping of fast reactions [6, 151]) and topological features of the model [153]. Simplified models hold the same benefits as smaller models in terms of computational effort and clarity. The reduction is usually achieved by lumping metabolites and/or reactions. Both cases are exemplified in Figure 5.3 C, where three lumped reactions and one lumped metabolite were introduced to the hypothetical glycolytic model. Merging reactions is more complex if the resulting lumped reaction skips intermediates that contribute to the biomass production. In the example, it would be wrong to assume that all acquired glucose molecules are converted to fructose-1-6-bisphosphate (fdp[c]) by the first lumped reaction *LR1*, as part of them are also used elsewhere (*e.g.* the pentose phosphate pathway and the direct

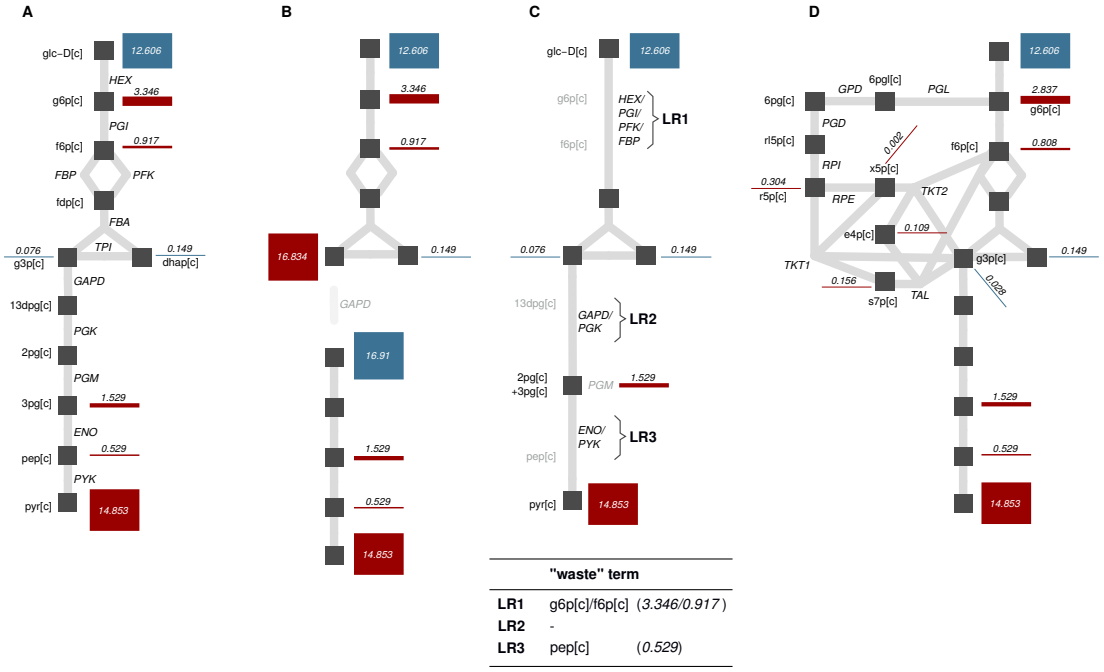


Figure 5.3: Calculating biomass leaks for split (**B**), simplified (**C**) and extended (**D**) models of yeast glycolysis (**A**). Bars show the magnitude of the leak flux of the corresponding dynamic metabolite (red - consuming, negative flux; blue - producing, positive flux) taken from [195]. The waste term in **C** quantifies the leaks that were omitted by lumping several reactions to the lumped reactions LR1-3. All leak fluxes are in mol/(mol biomass).
Abbreviations: glc-D[c] - glucose, g6p[c] - glucose-6-phosphate, f6p[c] - fructose-6-phosphate, fdp[c] - fructose-1,6-bisphosphate, dhap[c] - dihydroxyacetone-phosphate, g3p[c] - glyceraldehyde-3-phosphate, 13dpg[c] - bisphosphoglycerate, 2/3pg[c] - 2/3-phosphoglycerate, pep[c] - phosphoenolpyruvate, pep[c] - pyruvate, HEX - hexokinase, PGI - phosphoglucoseisomerase, PFK - phosphofructokinase, FBP - fructosebisphosphatase, FBA - fructosebisphosphate aldolase, TPI - triosephosphate isomerase, GAPD - glyceraldehyde-3-phosphate dehydrogenase, PGK - phosphoglucokinase, PGM - phoshoglycerate mutase, ENO - enolase, PYK - pyruvate kinase.

production of biomass). The subsequent *FBA* reaction would hence have to carry much less flux, which directly impacts on the estimation of its kinetic parameters. The backtracking workflow resolves this inaccuracy by introducing waste terms for each lumped reaction in the DDM as well as an amended stoichiometry. Importantly, the stoichiometry of the lumped reactions can be different between two distinct biological scenarios. In the example, *LR1* is consuming one glucose molecule but produces only 0.66 molecules of fdp[c] and a waste term as shown in the table in Figure 5.3 C. The occurrence of large waste terms for lumped DDM reactions may also help in the model design process as it hints at potentially important, but not yet included system features. For pooled species, the leak fluxes of all mapped metabolites belonging to the pool will be added up. The reactions inter-converting between pooled species are taken into account, if they have been mapped. This is especially important if they also turnover cofactors and ATP. ***Extending models*** is a joyful task when new data becomes available to improve the model. Backtracking allows to retain leak fluxes from the unchanged part of the model and consistently distributes the leak fluxes of the changed part to the new species and reactions. In the example, the pentose phosphate pathway was added to the model (Figure 5.3 D).

5.1.4 Results: Leak fluxes of the mycobacterial central carbon metabolism

Within the scope of this work, biomass backtracking was applied to calculate condition dependent leak fluxes from the central carbon metabolism in mycobacteria (formalized in the CCMtb model, *cf.* Chapter and Figure 6.1). The network includes the major carbon catabolic pathways glycolysis, pentose phosphate pathway, Krebs cycle, glyoxylate shunt, anaplerosis, as well as the methylcitrate and methylmalonyl pathways. A dummy respiratory chain is included to also account for ATP that is *de novo* synthesized via respiration.
As context, the most up-to-date GSM for *M. tuberculosis*, sMtb [159], was used with several small modifications, necessary to guarantee physiological function (summarized in Appendix A.9). The GSM contains an experimentally determined biomass composition, which will be the base for the calculated leak fluxes.

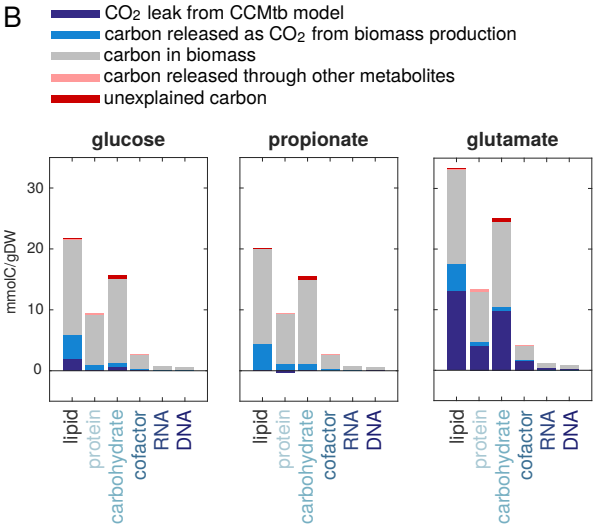
Leaks in three growth conditions

The leaks in the 3 experimental growth conditions (glucose, propionate, glutamate) were analyzed and exhibited a generally comparable distribution (Appendix A.10). In Figure 5.4 C, the distribution of leak fluxes required for the synthesis of the different biomass constituents for growth on glucose media is shown. For propionate and glutamate conditions, the distributions can be found in Appendix A.11.
The biomass of *M. tuberculosis* largely consists of complex lipids and carbohydrates constituting the thick cellular envelope that protects the bacteria against host induced killing (see Figure 4.1). Accordingly, almost 40 % of the taken up carbon is incorporated into the lipid fraction of the biomass, 30 % is channeled into the biosynthesis of complex carbohydrates. Lipids are predominantly build up from pyruvate, as well as acetyl- and propionyl-CoA. Flux towards the production of carbohydrates drains mainly from glycolysis. However, it becomes evident that both synthesis pathways, as well as the ones for RNA, DNA and protein, are highly complex, utilizing a variety of central carbon metabolism intermediates in vastly different stoichiometries. Backtracking allows to identify all contributions reliably based on the current knowledge on the entire metabolism of *M. tuberculosis*.

A

Rate (Cmmol/gDW)	glucose	propionate	glutamate
Uptake via C source	51.04	48.83	77.78
CO_2 leak from CCMtb	2.88	-0.25	29.22
CO_2 leak from biomass	5.72	6.75	5.98
Carbon in biomass	41.43	41.43	41.43
Release other metabolites	0.30	0.19	0.44
Unexplained carbon	0.70	0.70	0.70

B



C

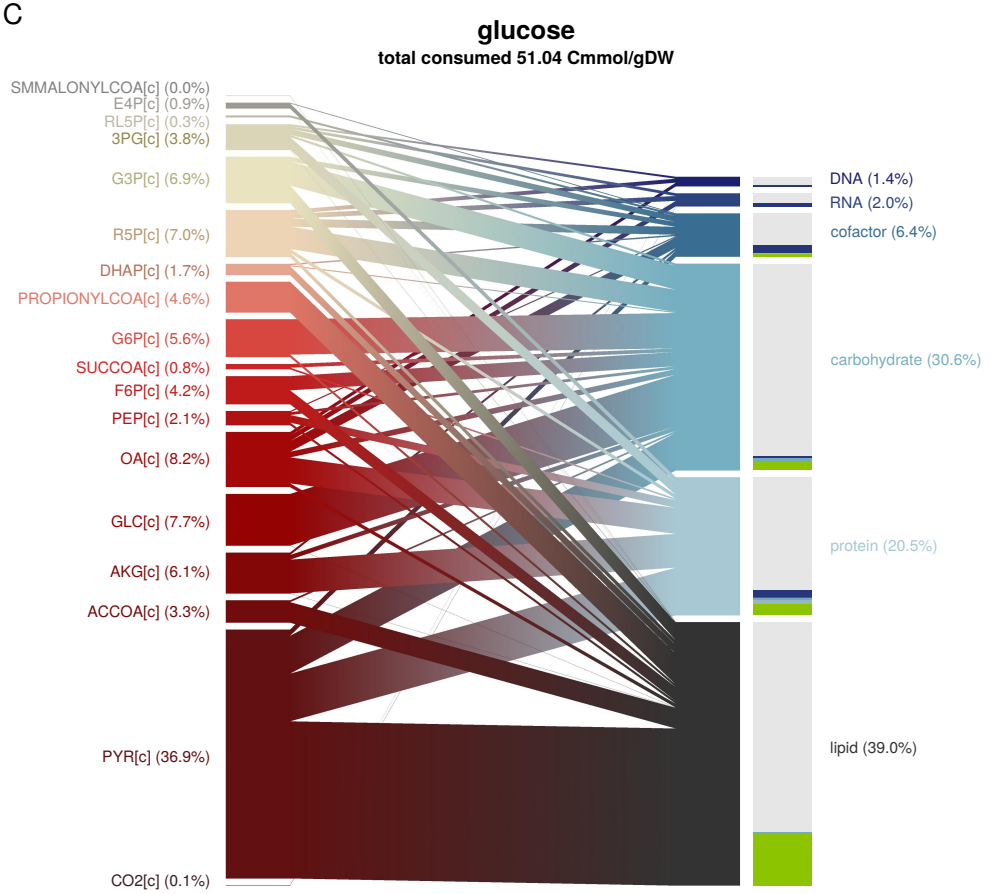


Figure 5.4: Biomass leaks in the three different growth conditions. **A** Summary of the fate of ingested carbon in the different growth conditions as derived from the backtracking. Negative values signify a resorption of carbon via CO_2 fixating reactions. **B** The same results as in A dissected into the requirements for subgroups of biomass components. **C** Distribution of leaks from CCMtb intermediates (left) to biomass constituents (right). In brackets, the contribution to the total amount of ingested carbon is indicated in Cmmol/gDW. Please note that only leaks with a negative sign are shown. Positive leaks imply compounds being released along the biosynthetic paths and re-entering the dynamic model. They are indicated here as colored, stacked bars on the biomass component's side (the green bar, for example, indicates released CO_2), and shown in detail in Appendix A.12. Abbreviations: Cmmol - mmol of carbon atoms, gDW - gram dry weight, species abbreviations are as in sMtb [159] and Appendix A.18.

Carbon conservation

As a quality and consistency check, the number of carbon atoms required for the production of each biomass component was analyzed. The results for the production of the entire biomass and single subgroups of biomass constituents are shown in Figure 5.4 A and B, respectively.
Theoretically, the sum of all carbon atoms of the dynamically modeled metabolites consumed in the leak should equal the number of carbon atoms per mole of biomass component, minus the amount that is released as carbon dioxide (CO_2) or other metabolites along the synthesis way. For CO_2, two different scenarios can occur: Either the CO_2 is released by reactions in the dynamically modeled network part (dark blue bars) or from distant biosynthetic reactions (light blue bars). The latter term would occur as a leak of positive sign for the CO_2 species of CCMtb (*i.e.* it is produced on the way to the biomass component and reinserted in the model), the former term would not, as it is already accounted for in the dynamic model. For the production of some biomass components additional secretion of other intermediates is required in sMtb (light red bars). This can be attributed to the minimalistic nature of the pFBA approach used in the backtracking, which sometimes favors unphysiological solutions with a smaller active network over more physiological ones which would include long range effects. This especially happens when only a single biomass constituent is backtracked (*cf.* Appendix A.13), as a number of network constraints are released in this optimization. If the used GSM is not fully consistent, unexplained carbon occurs as indication that the number of carbons in the atomic formula of the biomass component does not match the number of carbons incorporated ('Unexplained carbon' in Figure 5.4 A and B). This is the case for only a number of biomass components with difficult and variable composition (arabinogalactan/peptidoglycan and phosphatidylinositol mannosides), arguing for a high overall quality of the used GSM.
The propionate condition has the lowest overall loss to CO_2, the glutamate condition the highest. This is due to the fact that gluconeogenetic flux under glutamate conditions will have to pass through the reductive branch of the Krebs cycle. Hence, 1 of 5 acquired carbon atoms will be lost as CO_2 in the *kdh* or *kgd* reactions, which shows as a leak from the CCMtb model in Figure 5.4 B and Appendix A.14 (dark blue bars).

5.1.5 Discussion

Backtracking provides a solution to the problem of contextualizing detailed dynamic models of metabolic pathways. The condition dependent embedding allows a more realistic simulation of such pathways and their contribution to overall cellular functionality.
The overall mass balance, which – in contrast to isolated signaling models – has to be fulfilled by metabolic pathway models, usually complicates model calibration rather than easing it. Consistent and biologically meaningful leak fluxes ease this process. In this work, the calculated leak fluxes are essential for all other modeling approaches presented in the following Chapters. As shown above, the backtracking can also support an easy quality control of the used GSM, on whose accuracy and experimental backing the calculated leak fluxes will rely.
Biomass backtracking can also be seen as a means of large scale data integration. Data from omics measurements are frequently used to restrict the feasible flux space in sophisticated FBA methods [38, 42, 76, 98, 103, 210]. Leak fluxes calculated from such experimentally constraint flux distributions will transmit the information to the pathway of interest – as some kind of simple transfer function that can be refined as needed and as the data situation allows. This might also have applications in the integration of patient data for personalized medicine.
Another focus of application is the iterative development and improvement of models as is

required for modular or whole-cell modeling [28, 86, 94]. Modified models will still need to match the overall context and, moreover, interfaces between modules have to be consistent. Backtracking provides a reliable and reproducible way to ensure these criteria.

Limitations

Backtracking in different nutritional conditions cannot account for changes in the actual composition of the biomass, which is very likely to change between different growth conditions: It has been shown for the lipid content during different growth phases [169] or different growth rates in the chemostat [20]. The single component backtracking, provides a way to generate distinct reactions for the synthesis of each component. If individual, and potentially regulated reaction kinetics are defined for each of those reactions, the composition can be easily adapted in accordance with the biological context.

Unfortunately, the plasticity of biomass production pathways if one precursor is missing cannot be represented by the model with fixed biomass drains. The production of the biomass will cease as soon as one of the precursors is no longer available. A next step would be to develop new methods to more precisely couple dynamic simulations with constraint models of metabolism and simulate them together.

5.2 Extending driving force optimality to the network context

While backtracking defines sensitive boundary fluxes for metabolic models, it is solely based on the steady state assumption of the underlying network and on minimizing the overall enzyme production cost by applying pFBA. However, these are only some of the conceivable constraints that cellular metabolic networks could be subjected to. An inevitable constraint will be the fulfillment of energy balance (*i.e.* thermodynamic consistency), while others could include building up defensive mechanisms, like a thick cell wall as in mycobacteria [44], nutrient storages for potential harsher times [183] or maintaining transmembrane ion gradients [208].
Within the central carbon metabolism, a good coverage of measured metabolite concentrations and thermodynamic reaction enthalpies can be achieved, such that they can be incorporated in order to get a deeper understanding of the flux distributions within this part of the metabolic system. This can refine the findings based on the thermodynamic characterization of the network by the reversibility index (Section 4.2.2) by additionally incorporating the network topology.
For this task, a number of FBA based methods are available, as outlined below. In addition, this work focuses on applying a measure of optimality to the thermodynamic landscape, which is new to the world of network thermodynamics. Thereby, existing methods are combined to a novel setup assuring mass and energy balance while striving for thermodynamic optimality.

5.2.1 A network approach to maximum-minimum-driving forces

Previous methods to link mass and energy balances

A number of methods are already available to combine stoichiometric network models with thermodynamic constraints, examining consistency and partly also optimality of flux distributions and metabolomic datasets.
Network embedded thermodynamics [98, 212] strive to calculate boundaries for physiologically possible $\Delta_r G'$ values or metabolite concentrations given a set flux distribution in a network, based on the consistency of reaction direction and the sign of $-\Delta_r G'$. Flux directions are thereby inferred from experimental flux measurements, flux balance analysis or can be left undefined where unknown. This last feature, however, while reducing the number of assumptions, does not guarantee a feasible steady state of the network.
The inverse approach of defining feasible flux directions from thermodynamic constraints was formalized by Beard *et al.* (***Energy Balance Analysis*** [14, 15]) and systematically applied to the then available genome-scale network of *E. coli* by Henry and coworkers (***Thermodynamic FBA/TMFA*** [71]). They thereby applied default bounds on metabolite concentrations as well as, after an unfeasible system resulting from the initial approach, included uncertainties in $\Delta_r G^0$.
As a combination of the above methods, Hoppe *et al.* directly included information on metabolite concentrations in the thermodynamically consistent flux calculation (***Thermodynamic Realizability*** [76]). Also here, measurement errors of the $\Delta_r G^0$ were accounted for. Additional bounds corresponding to measured metabolite concentrations were elegantly implemented by partly nonlinear penalty functions. As a computational drawback, the resulting optimization problem has a nonlinear scoring function, which is difficult to solve numerically.
On the pathway level, where the direction of the flux for all reactions is known, ***pathway thermodynamics*** [134] can be applied. Thermodynamic principles are here applied to transition from a feasible, pathway-wide thermodynamic gradient to an optimal one. Optimality is thereby formulated as the least feasible reaction in the pathway reaching its maximum possible driving force (*i.e.* the inverse of $\Delta_r G'$) in the pathway context. Additionally to this new optimality thought, also the optimization problem formulates as a linear, easily solvable one.

This approach

The approach presented here combines the pathway based calculation of the maximum-minimum driving force (MDF) via pathway thermodynamics [134] with the network oriented TMFA [71] framework. This allows to calculate a set of thermodynamically optimal metabolite concentrations and flux directions for the central carbon metabolism in mycobacteria.

5.2.2 Mathematical formulation

The desired approach can be formalized as a linear optimization problem with thermodynamic and mass balance constraints on the flux distribution. To optimize for a favorable thermodynamic landscape, the MDF is introduced to the optimization as a common upper bound on the Gibbs energies of active fluxes. Its minimization will force all $\Delta_r G'$ to decrease in order to achieve a maximally steep thermodynamic gradient for all reactions. The required constraint equations and possible objective function for the implementation are described in the following.

Set of Constraints

$\mathbf{N}' \in \mathbb{R}^{m \times r}$ is defined as the stoichiometric matrix of the irreversible version of the metabolic network in question, with r as the number of reactions and m as the number of metabolites. To obtain this version, all reversible reactions are split in one reaction for each direction. For a subset of $\mathbf{r}_t$ reactions (excluding transport and biomass reactions) thermodynamic constraints were included. $\mathbf{N}'_t \in \mathbb{R}^{m \times r_t}$ is the corresponding part of $\mathbf{N}'$.
The vector of variables to be optimized $\mathbf{d}$ includes the metabolic fluxes $\mathbf{v} \in \mathbb{R}^r$, the logarithmic metabolite concentrations $\mathbf{x} \in \mathbb{R}^m$, binary variables $\mathbf{u} \in \mathbb{R}^r$ and $\mathbf{q}, \mathbf{z} \in \mathbb{R}^{r_t}$ corresponding to pointers for positive flux, positive $\Delta_r G'$ and a flux direction violating the thermodynamic gradient, respectively, the violating flux values $\mathbf{w} \in \mathbb{R}^{1 \times r_t}$, the reaction Gibbs energies $\mathbf{\Delta}_r \mathbf{G}' \in \mathbb{R}^{r_t}$ (normalized by RT), the standard Gibbs energies $\mathbf{\Delta}_r \mathbf{G}^0 \in \mathbb{R}^{r_t}$ and the maximum-minimum-driving force MDF, which is the target of the optimization. $\Delta_r G'_{max}$ and V_{max} are large numbers outside the range of possible $\Delta_r G'$ and flux values, respectively.
A linear optimization problem as described in Section 2.19 with the following set of constraints is solved:

$$\mathbf{N}' \cdot \mathbf{v} = 0 \qquad \text{mass balance (steady state) constraints} \qquad (5.1)$$

$$\mathbf{N}_t'^{\mathrm{T}} \cdot \mathbf{x} - \mathbf{\Delta}_r \mathbf{G}' - \mathbf{\Delta}_r \mathbf{G}^0 = 0 \qquad \Delta_r G' \text{ according to concentrations in } \mathbf{x} \qquad (5.2)$$

$$0 \leq \mathbf{v} \leq V_{max} \qquad \text{lower and upper bounds on the flux} \qquad (5.3)$$

$$\mathbf{x}_{lb} \leq \mathbf{x} \leq \mathbf{x}_{lb} \qquad \text{data based lower and upper bounds on the metabolite concentrations} \qquad (5.4)$$

$$\mathbf{\Delta}_r \mathbf{G}^0_{lb} \leq \mathbf{\Delta}_r \mathbf{G}^0 \leq \mathbf{\Delta}_r \mathbf{G}^0_{ub} \qquad \text{data based lower and upper bounds on } \Delta_r G^0 \qquad (5.5)$$

$$\mathbf{v} - \mathbf{u} \cdot V_{max} \leq 0 \qquad \text{pointer } \mathbf{u} \text{ for positive flux} \qquad (5.6)$$

$$\mathbf{\Delta}_r \mathbf{G}' - \mathbf{q} \cdot \Delta_r G'_{max} < 0 \; (\leq -\varepsilon) \qquad \text{pointer } \mathbf{q} \text{ for positive } \Delta_r G' \qquad (5.7)$$

$$\mathbf{u} + \mathbf{q} - \mathbf{z} \leq 1 \qquad \text{violating flux pointer } \mathbf{z} \qquad (5.8)$$

$$\mathbf{v} + \mathbf{z} \cdot V_{max} - \mathbf{w} \leq V_{max} \qquad \text{flux value of violating reactions} \qquad (5.9)$$

$$\mathbf{w} \geq 0 \qquad \text{lower bound for violating flux} \qquad (5.10)$$

$$\mathbf{\Delta}_r \mathbf{G}' + \mathbf{u} \cdot \Delta_r G'_{max} - MDF \leq \Delta_r G'_{max} \qquad \text{MDF constraint on all active fluxes} \qquad (5.11)$$

The system of constraints can be interpreted as follows: To ensure mass conservation the sum of all changing fluxes **v** for each metabolite in **x** has to be 0 (constraint 5.1). As the biomass leaks are included as one combined biomass reaction with the respective stoichiometric coefficients, this will also guarantee biomass production in the right proportions. The flux through this reaction was fixed to arrive at a division time of 38 h (*cf.* Table 4.2).
The dependency of $\mathbf{\Delta}_r\mathbf{G}'$ on $\mathbf{\Delta}_r\mathbf{G}^0$ and the logarithmic metabolite concentrations **x** is implemented in constraint 5.2. Note that here also the $\mathbf{\Delta}_r\mathbf{G}^0$ are variables rather than fixed values. The constraints 5.3-5.5 implement bounds on the fluxes, metabolite concentrations and on $\mathbf{\Delta}_r\mathbf{G}^0$ (based on the measured data: 2 standard deviations (*sd*) for the metabolite data, 3*sd* for the standard Gibbs energies). Lumped observables, *i.e.* measurements of the sum of two or more metabolites, are resolved to allowing all summand metabolites to vary within the full range of the summed concentration, resulting a maximum 2-fold error. If the summation would be implemented directly the problem would become nonlinear as the space of logarithmic concentrations is searched.
Constraint 5.6 introduces the boolean pointer **u** for positive flux, which is forced to be 1 for all positive fluxes in order to fulfill the constraint equation. The same holds for pointer **q** and the $\mathbf{\Delta}_r\mathbf{G}'$ in constraint 5.7. This constraint has to be a strict inequality, as also for the equilibrium at $\Delta_r G' = 0$, no net flux can occur. As the common linear programming solvers do not account for strict inequalities, it was resolved to a non-strict inequality by the introduction of $\varepsilon > 0$.
In the original TMFA, constraint 5.8 would read as $\mathbf{u} + \mathbf{q} \leq 1$, assuring that whenever a positive flux forces its respective pointer **u** to be 1, the corresponding Gibbs energy pointer **q** needs to be 0, encoding a negative $\Delta_r G'$. However, if the problem with the given constraints and data is infeasible, as was the case here, this constraint has to be modified to identify exhausted constraints. On the way to a feasible solution, the pointer **z** was introduced to soften this constraint. All fluxes violating the relation between flux direction and the sign of $\mathbf{\Delta}_r\mathbf{G}'$ will force a 1 in the respective position of **z**. Along with constraint 5.9 that introduces the vector **w** containing the actual values of all infeasible fluxes, this allows for a systematic reduction of infeasibilities as described below. Constraint 5.10 corresponds to the lower flux bound for the violating fluxes, in order to guarantee again a irreversible network.
Finally, the MDF, as an upper bound on $\mathbf{\Delta}_r\mathbf{G}'$ for flux carrying reactions marked in **u**, is implemented in constraint 5.11. In contrast to the original publication by Noor *et al.* [134], where the MDF is the lower bound for the driving force -$\Delta_r G'$, the inverse notation is used to allow for a common optimization sense ("minimization") if the MDF is combined with other objectives.
As in the original TMFA, the resulting problem formulates as a mixed zero-one linear programming, as some of the variables are booleans whereas some are continuous. The solution of this type of problems is computationally expensive and requires long computation times.

Objective function

The objective function was intended to represent both optimum physiological flux distribution as well as a high driving force. The first aspect is commonly implemented by parsimonious FBA, while for the latter aspect, the optimization of the MDF as described by Noor *et al.* is employed. The overall objective then reads as:

$$\min_{\mathbf{x},\mathbf{v},\mathbf{\Delta}_r\mathbf{G}^0,MDF,\mathbf{u},\mathbf{y},\mathbf{z},\mathbf{w}} (MDF + w_{pFBA} \sum_{i=1}^{r} v_i), \tag{5.12}$$

with w_{pFBA} as weight with which the parsimonious FBA contributes to the optimum solution.
However, this objective function gives biologically meaningful results only if the search space

Table 5.1: Summary of the feasible solutions of the described optimization problem for the 4 experimental conditions.

	Glucose, early	**Glucose**, late	**Glutamate**, late	**Propionate**, late
runtime (s)	0.99	2.47	108.62	37.74
pFBA weight w_{pFBA}	0.01	0.001	0.001	0.001
Max-min driving force (MDF, kJ/mol)	**1.59**	**3.23**	**1.24**	**0.046**

contains a feasible solution. When this is not the case, the bounds have to be systematically relaxed to arrive at a feasible solution, for which the following other objectives were used: (i) Minimizing the number of all infeasible reactions ($\min \sum \mathbf{z}$) and (ii) minimizing the sum of fluxes for all infeasible reactions ($\min \sum \mathbf{w}$). Both approaches help to identify critical, infeasible reactions, as long as the system is not thermodynamically feasible, but can include flux loops and non-physiological solutions (*e.g.* futile cycles). For this reason, those objectives were abandoned as soon as the system was feasible.

Iterative MDF

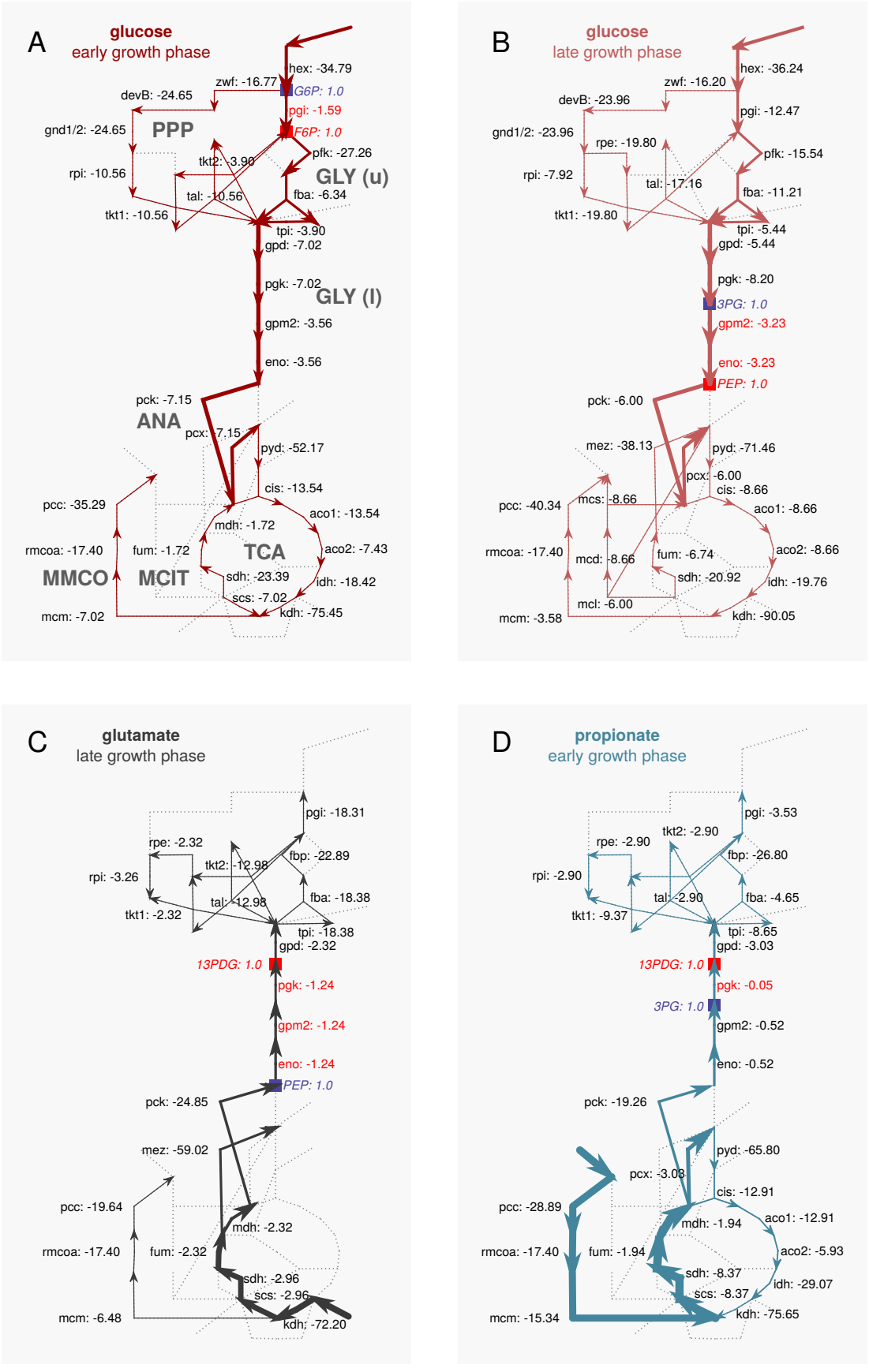
As the MDF only acts as an upper bound for $\mathbf{\Delta}_r\mathbf{G}'$, it is usually possible to lower the remaining reactions' $\Delta_r G'$ further than in the found solution to the MDF problem, to arrive at a more beneficial thermodynamic landscape. For this, shadow prices of the optimization result can be exploited to identify limiting constraints. Here, the reactions employing only metabolites with the highest shadow prices, *i.e.* with a $\Delta_r G'$ equal to the found MDF value, were fixed to this found value. With the given optimum flux distribution, the MDF for the remaining reactions was optimized again. This process was repeated until all reactions were fixed, arriving at the best network-wide driving force distribution.

5.2.3 Results: Resolving infeasibilities

To arrive at a mass balanced system with an optimal thermodynamic landscape, the central metabolic network along with the corresponding metabolite measurements from four different cultivation conditions were iteratively analyzed.

No feasible solution with the initial bounds on $\Delta_r G'$ and metabolite concentrations could be found, except for the late exponential growth on glucose. A combination of different objective functions that were used to identify infeasible reactions is summarized in Appendix A.16, employing the described markers $\mathbf{z}$ and flux values $\mathbf{w}$ for infeasible reactions. For the identified reactions the corresponding bounds on the metabolite concentrations were expanded by 4 *sd* (based on the experimental data). This was the case for the *tpi* reaction on early glucose and the *pgk* reaction on glutamate and propionate.

Figure 5.5: Thermodynamically optimal, mass balanced flux distributions in the central metabolic network. For each measured experimental condition (**A** Glucose, early growth phase, **B** Glucose, late growth phase, **C** Glutamate, late growth phase, **D** Propionate, early growth phase) flux distributions are shown, where the width of the arrows represents the strength of the flux. Please not that for clarity biomass leaks are not shown, but drain and feed into the network in the calculations (*e.g.* fumarate feed explains *sdh* flux direction). In **A** the involved pathways are labeled in gray as: PPP - pentose phosphate pathway, GLY - glycolysis (upper, u and lower, l), ANA - anaplerotic/gluconeogenetic reactions, TCA - Krebs cycle, MCIT - methylcitrate pathway, MMCO - methylmalonyl pathway. Each reaction is labeled with its $\Delta_r G'$ in kJ/mol after the iterative optimization, the bottleneck reactions are shown in red. For detailed information on the shown reactions see Appendices A.20 and A.21 . Metabolites with the highest shadow prices are highlighted as boxes (red - negative shadow price, blue - positive shadow price, the price itself is printed behind the metabolites' names). Dashed lines indicate reactions that do not carry flux.


A
glucose
early growth phase
hex: -34.79
zwf: -16.77
devB: -24.65
G6P: 1.0
pgi: -1.59
gnd1/2: -24.65
PPP
F6P: 1.0
pfk: -27.26
tkt2: -3.90
rpi: -10.56
GLY (u)
fba: -6.34
tal: -10.56
tkt1: -10.56
tpi: -3.90
gpd: -7.02
pgk: -7.02
GLY (l)
gpm2: -3.56
eno: -3.56
pck: -7.15
ANA
pcx: -7.15
pyd: -52.17
cis: -13.54
pcc: -35.29
aco1: -13.54
mdh: -1.72
rmcoa: -17.40
fum: -1.72
TCA
aco2: -7.43
MMCO
MCIT
sdh: -23.39
idh: -18.42
scs: -7.02
mcm: -7.02
kdh: -75.45
B
glucose
late growth phase
hex: -36.24
zwf: -16.20
devB: -23.96
pgi: -12.47
gnd1/2: -23.96
rpe: -19.80
pfk: -15.54
rpi: -7.92
tal: -17.16
fba: -11.21
tkt1: -19.80
tpi: -5.44
gpd: -5.44
pgk: -8.20
3PG: 1.0
gpm2: -3.23
eno: -3.23
PEP: 1.0
pck: -6.00
mez: -38.13
pyd: -71.46
pcx: -6.00
cis: -8.66
pcc: -40.34
mcs: -8.66
aco1: -8.66
rmcoa: -17.40
mcd: -8.66
fum: -6.74
aco2: -8.66
sdh: -20.92
idh: -19.76
mcl: -6.00
mcm: -3.58
kdh: -90.05
C
glutamate
late growth phase
pgi: -18.31
rpe: -2.32
tkt2: -12.98
fbp: -22.89
rpi: -3.26
fba: -18.38
tal: -12.98
tkt1: -2.32
tpi: -18.38
gpd: -2.32
13PDG: 1.0
pgk: -1.24
gpm2: -1.24
eno: -1.24
PEP: 1.0
pck: -24.85
mez: -59.02
pcc: -19.64
mdh: -2.32
rmcoa: -17.40
fum: -2.32
sdh: -2.96
scs: -2.96
mcm: -6.48
kdh: -72.20
D
propionate
early growth phase
pgi: -3.53
tkt2: -2.90
rpe: -2.90
fbp: -26.80
rpi: -2.90
tal: -2.90
fba: -4.65
tkt1: -9.37
tpi: -8.65
gpd: -3.03
13PDG: 1.0
pgk: -0.05
3PG: 1.0
gpm2: -0.52
eno: -0.52
pck: -19.26
pyd: -65.80
pcx: -3.08
cis: -12.91
pcc: -28.89
mdh: -1.94
aco1: -12.91
rmcoa: -17.40
fum: -1.94
aco2: -5.93
sdh: -8.37
idh: -29.07
scs: -8.37
mcm: -15.34
kdh: -75.65

Also after this release, which allowed for a fully feasible flux distribution, these reactions remained part of the thermodynamic bottleneck in the respective conditions. The MDF reaches the lowest value on propionate, comparable higher values for glutamate and early exponential glucose conditions, and the best value for the late exponential glucose condition (summarized in Table 5.1, the found optimum metabolite concentrations along with their bounds are depicted in Appendix A.17).

5.2.4 Results: Glycolysis as thermodynamic bottleneck

The found optimum flux and thermodynamic gradient distribution is mapped to the central metabolic network considered in Figure 5.5 and summarized in Figure 5.6. In all examined conditions, glycolysis reveals itself as thermodynamic bottleneck (reactions are indicated in red in Figure 5.5, with corresponding optimal $\Delta_r G'$equaling the MDF). From early to late exponential growth phase on glucose, the thermodynamic bottleneck slides from upper to lower glycolysis, thereby always marking non ATP coupled reactions. For the gluconeogenetic carbon sources glutamate and propionate, lower glycolysis including the *pgk* reaction is the most inefficient part, being barely feasible on propionate ($\Delta_r G'$ = -0.05 kJ/mol, J^-/J^+ = 0.98).
For all conditions, the Krebs cycle runs in the generic (oxidative) direction, most likely due to the extremely low $\Delta_r G^0$ of the *kdh* reaction. This argues for favorable redox conditions, where a reductive reversal of the left half of the cycle is not required [204]. If glutamate as carbon source is supplied, the right half of the Krebs cycle is not required, as no biomass components drain from there (*cf.* Figure 5.4 C and Appendix A.11).
For the production of virulence relevant lipids, intermediates of the methylmalonyl and methylcitrate pathways are required. In all conditions, their production via the methylmalonyl pathway

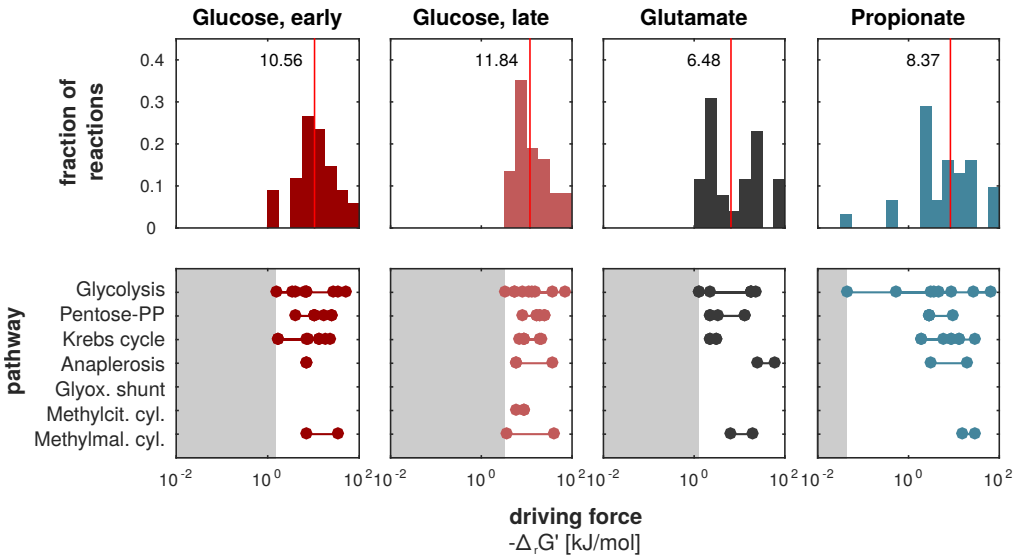


Figure 5.6: Network wide summary of the found optimum solution. The upper panel shows a histogram for all reactions' driving forces (-$\Delta_r G'$ in kJ/mol) for the 4 experimental conditions, with the median marked as red line and corresponding value. The lower panel shows the occurring driving force values dissected in the single pathways considered in the network (dots), lines depict the range for each pathway. The gray areas illustrate the found lowest driving force.

is preferred. The methylcitrate cycle only becomes active in the late growth phase on glucose, a condition for which its intermediate metabolites were not quantified. As explained before (see Figure 4.4), the number of measured metabolites influences the definition of the thermodynamic gradient and the reaction reversibility, such that this result could be an artifact.
The PEP-pyruvate-oxaloacetate node, which distributes carbon between gluconeogenesis/ anaplerosis and the Krebs cycle [172], shows non-intuitive behavior, especially for the glucose conditions, where the *pyk* reaction, the second last step in glycolysis, is bypassed. In the late growth conditions, several reactions share the title of the least feasible reaction. This has been described earlier as the concept of distributed bottlenecks, in which a set of reactions, rather than only one, becomes thermodynamically limiting [118].
While the MDF is lowest on propionate, the overall distribution of $\Delta_r G'$ in this condition is more favorable compared to growth on glutamate (Figure 5.6). In this comparison, also early and late glucose conditions show a more comparable gradient distribution that also has a higher median compared to the gluconeogenetic carbon sources.

5.2.5 Discussion

Towards thermodynamic network optimality

In the described approach, the central carbon metabolic network of *M. bovis* BCG was evaluated for both thermodynamic consistency and optimality. While there is room for improvement, a first endeavor to apply thermodynamic optimality principles on purely stoichiometric networks is presented. Glycolysis is identified as critical bottleneck in all tested conditions. It can be argued that due to their reversible nature, glycolytic reactions cannot be proficient in one direction, especially in an organism relying on fast rerouting of its metabolism in changing environments. Not surprisingly, the thermodynamic landscape is more beneficial on the high energy compound glucose than on the gluconeogenetic carbon sources propionate and glutamate (Figure 5.6).

Bias of measured metabolites

As already shown for the calculation of the reversibility index, a strong bias by the number of measured educts for a reaction becomes evident. For example could the striking differences between the two glycolytic conditions potentially be an artifact of missing metabolite measurements, that allow the MCIT pathway to be used in reverse direction (Figure 5.5 B), thereby rerouting many fluxes in the right half of the Krebs cycle. Also, the additional measurement of F6P in the early growth condition tightly constrains the *pgi* reaction, even defining a new overall bottleneck compared to the late growth condition, where this measurement is lacking.
It can hence be concluded that for understanding the optimal thermodynamic landscape as opposed to a merely feasible one, quantitative metabolomics measurements are indispensable.

Further use

For the work presented in the following Chapters, the found infeasibilities in the metabolite data set will be considered. The results will be used to relax the bounds on metabolite concentrations where specifically necessary to guarantee thermodynamic feasibility in later applications of the model and data. The allowable changes are summarized in Table 5.2.

Condition	**Metabolites** on which bounds were relaxed
Glucose, early	-
Glucose, late	G3P, DHAP
Glutamate, late	13PDG, 3PG
Propionate, early	13PDG, 3PG, PEP

Table 5.2: Metabolites for which bounds based on the measured metabolite concentrations were released to allow for thermodynamic network feasibility.

Limitations and possible improvements

While combining network derived flux feasibility with thermodynamic optimality has the potential to improve understanding of cellular metabolism, the method has some practical limitations discussed in this Section along with possible ways to resolve them in future works.

A major limitation is the computationally hard formulation as a mixed zero-one linear programming problem, which is not efficiently solvable with the currently available linear programming solvers. The nature of the problem arises from the binary 'choosing' of different fluxes to be active or not in the given flux distribution in addition to the maximization of the driving force, which is in itself an easily solvable linear problem.

A complementary approach could be the calculation of thermodynamically feasible elementary flux modes as described by Gerstl *et al.* (tEFMA [65]). These paths trough a network include all possible ways to fulfill the flux balance (and hence also the appropriate biomass formation) in a minimal, non-splittable fashion. If all elementary flux modes of the network are known, the original linear MDF optimization could be easily used to thermodynamically optimize each of them and select the best one, representing a global solution that is minimal in flux expenditure and maximal in the driving force distribution. However, the problem of finding all elementary flux modes in the first place, is again computationally expensive [213].

Even with this potential global solution, a relation between the actual extent of the flux and the thermodynamic feasibility is lacking in all of the above methods. From an overall optimality point of view, it might not be important to improve the thermodynamic gradient for all reactions, but only for those required to carry a large flux to attain a large effect on the total enzyme costs. However, this cost does not only depend on the thermodynamic landscape, but also on the mass and the saturation state of the enzymes as well as on active regulatory effects. To describe these dependencies, detailed information on the kinetic properties of the involved enzymes are required as will be described in the following Chapters.

6. CCMtb: Modeling central carbon metabolism in mycobacteria

Where are regulatory points in the mycobacterial central carbon metabolism?

The flux through a metabolic network is limited, but not fully defined, by the thermodynamic landscape characterized in the previous Chapter. As stated before, the flux also directly depends on the amount of enzyme available. However, even if the enzyme concentrations are known from measurements, kinetic and regulatory effects have large impact on the flux distribution. While the previous Chapter focused on thermodynamics, the following Chapter will employ enzyme concentrations and kinetic descriptions of each reaction in the central carbon metabolism. The compilation of a detailed dynamic model will be described, which includes these two factors of intracellular flux regulation. Subsequently, the model will be used to identify potential regulatory points for the transition between the experimental carbon sources in a new model and data driven optimization approach.

6.1 Compiling a detailed model of central carbon metabolism

Model Topology

The model includes the major carbon catabolic pathways in mycobacteria: Glycolysis, pentose phosphate pathway, Krebs cycle, glyoxylate shunt, the propionate catabolic methylmalonyl and methylcitrate pathways as well as anaplerotic/gluconeogenetic reactions. Topological information was collected from literature as well as the Tuberculist [101], KEGG [84] and BRENDA [175] databases. All reactions are summarized in Appendix A.21 and depicted in Figure 6.1. The same model topology was also used for backtracking and the thermodynamic considerations in Chapter 5. As stated in the Methods Section 3.3, metabolomics data were mapped to the network via observables (Appendix A.19).

The genes associated with each reaction were collected as described in the Methods Section 3.2 and are summarized in Appendix A.20. Each enzymatic reaction was modeled with simple convenience kinetics [104], as an automated way to set up kinetic laws with varying stoichiometry. Please note, that the purely kinetic form of the rate law is used here, not the thermodynamic one (for equations see Appendix A.22).

A condition dependent biomass reaction with a stoichiometry based on the backtracking results was included. Backtracking only provides the stoichiometric coefficients for the biomass reaction, the kinetic description in the dynamic model has to be defined additionally. Here, its kinetics are described by a fixed rate r_{bio}. The backtracked leak fluxes are given in mmol/gDW. To convert them to the μM units representing intracellular concentrations that are used in the further analyses, they need to be normalized to the dry weight and the volume of one cell (*cf.* Table 4.1). The resulting conversion factor of $432 \cdot 10^3 \frac{\mu\text{M}}{\text{mmol/gDW}}$ would yield high stoichiometric coefficients for the biomass reaction, which lead to numeric instabilities in the simulation.

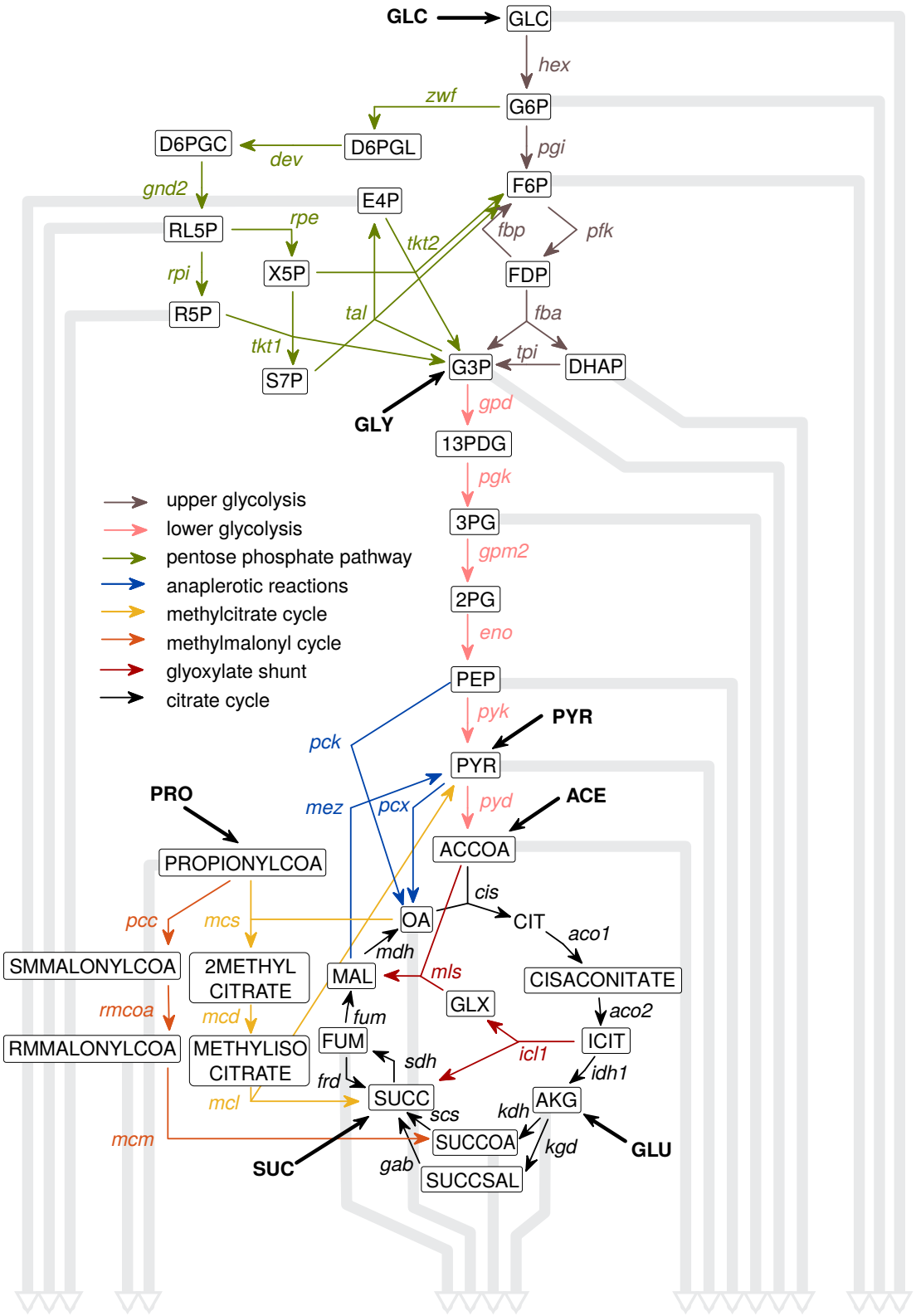

GLC
GLC
hex
G6P
zwf
D6PGL
D6PGC
dev
pgi
F6P
gnd2
E4P
RL5P
rpe
fbp
pfk
tkt2
rpi
X5P
FDP
R5P
tal
fba
tkt1
tpi
G3P
DHAP
S7P
GLY
gpd
13PDG
pgk
upper glycolysis
lower glycolysis
pentose phosphate pathway
anaplerotic reactions
methylcitrate cycle
methylmalonyl cycle
glyoxylate shunt
citrate cycle
3PG
gpm2
2PG
eno
PEP
pyk
PYR
pck
PYR
PRO
mez
pcx
pyd
ACE
ACCOA
PROPIONYLCOA
cis
pcc
mcs
OA
CIT
aco1
mdh
SMMALONYLCOA
2METHYL CITRATE
MAL
mls
CISACONITATE
rmcoa
mcd
GLX
aco2
fum
RMMALONYLCOA
METHYLISO CITRATE
FUM
ICIT
icl1
sdh
idh1
frd
mcl
SUCC
AKG
scs
kdh
mcm
SUCCOA
kgd
GLU
SUC
gab
SUCCSAL

Figure 6.1: Summary of the enzymatic reactions contained in the CCMtb model. Reactions are shown as colored arrows, according to the pathway they are associated with and labeled with the model reaction name (Appendix A.20 and A.21). All reactions are reversible, reaction arrows are depicted in their forward direction as implemented in the model. Metabolites are shown in capital letters, their names are listed in Appendix A.18 along with further annotations. Thick black arrows mark entry points of the experimental carbon sources to the network. Gray arrows denote all occurring biomass leaks (width not to scale).

As the model is simulated in seconds, this can be avoided by including the conversion from hours to seconds (1/3600) in the calculation, reducing the conversion factor to $120\frac{\mu\text{M}}{\text{mmol/gDW}}$. Accordingly, the rate of biomass production in the model will correspond to the value of r_g in 1/h, also here resulting in numerically more favorable values (*e.g.* for a division time of t_d of 38 h the rate will equal 0.0182 1/s as opposed to $5.07 \cdot 10^{-6}$ 1/s.).
Some metabolites were allowed to be secreted, based on the BCG reference experiment (succinate, acetate, isocitrate, see Appendix A.2) or on general biological behavior (phosphate, water, carbon dioxide). The latter group was also allowed to be taken up from the media. Uptake rates were modeled as fluxes with a fixed rate r_{up}, depending on the nutritional condition. Secretion reactions are equipped with Michaelis-Menten kinetics.
To convert between energy and redox species, a reduced version of the respiratory chain was implemented as described in detail below. The network includes reactions that employ ATP and GTP as well as NADH and NADPH. For thermodynamic consistency, the metabolites were kept in the reaction, even though their turnover is drastically affected by reactions outside the modeled network. To avoid over-constraining the model, two converting dummy reactions Rresp_trd and Rresp_ndk were added for equilibrating NADH and NADPH as well as ATP and GTP pools, respectively.

Respiratory Chain

The respiratory chain in *M. tuberculosis* is not fully understood yet. It plays a role in the pathogenic context, for example in the adaptation to hypoxia [64, 204]. *M. tuberculosis* encodes many different genes that can participate in respiration (for a summary see Cook *et al.* [41]). Figure 6.2 shows the general structure as it is also implemented in the modified version of the genome-scale metabolic model sMtb [159] which was used in the backtracking before. The reactions can be summarized as:

NUO	$5\ H^+$	+	NADH	+	MQ	$\rightarrow$	$4\ H_e^+$	+	NAD^+	+	MQH_2,
SDH	SUCC	+	MQ	+	$2\ H_e^+$	$\rightarrow$	MQH_2	+	FUM,		
CYDB	$6\ H^+$	+	O_2	+	$2\ MQH_2$	$\rightarrow$	$6\ H_e^+$	+	$2\ H_2O$	+	2 MQ,
ATPs	$3\ H_e^+$	+	ADP	+	Pi	$\rightarrow$	ATP	+	H_2O	+	$3\ H^+$,

with the abbreviations as in Figure 6.2. Electrons are carried by menaquinols (MQH_2) in the mycobacterial cell membrane.
CCMtb includes a simplified version of the above respiratory chain. Simplifications were made to ensure matching levels of detail between data and model and to avoid the introduction of compartments. However, the simplified version should keep the overall energetic balance as well as contributions of single pathways, as for example the Krebs cycle contribution via succinate dehydrogenase (SDH). It was also intended to conserve the intracellular metabolite species to guarantee comparability to the measured metabolite concentrations.

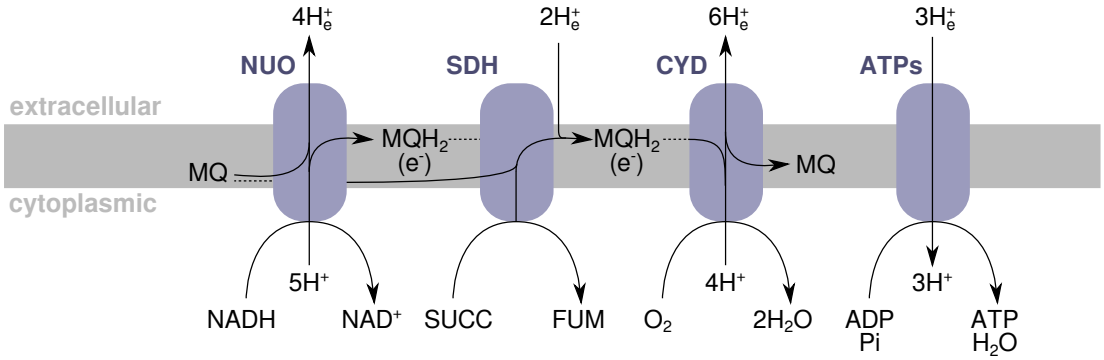


Figure 6.2: Schematic representation of the respiratory chain in *M. tuberculosis* based on [24, 41]. Enzyme complexes are shown in blue and named as in the sMtb model, the gray area represents the plasmamembrane.
Abbreviations: NUO - NADH:menaquinone oxido-reductase, SDH - succinate:menaquinone oxidoreductase, CYD - cytochrome oxidase, ATPs - F1F0-type ATP synthase, MQ - menaquinone, MQH_2 - menaquinol, SUCC -Succinate, FUM - Fumarate, H_e^+ - extracellular protons, H^+ - intracellular protons, Pi - inorganic phosphate.

As the reactions are encoded by a pleiotropy of different genes [24, 41], whose contribution to respiration is not fully clarified, no gene association was included for the respiratory chain reactions, *i.e.* the enzyme amount is implemented in the kinetic equation as v^{max} (see Appendix A.22). Nevertheless, the thermodynamic gradient can be accounted for by calculating $\Delta_r G^0$ for the overall reaction from the eQuilibrator database [58].
The above reactions were hence simplified to three lumped reactions as:

Rresp_nadh	NADH	+	2.3 MQ						$\leftrightarrow$	NAD^+	+	2.3 MQH_2,
Rresp_atps	H_2O	+	ADP	+	Pi	+	MQH_2		$\leftrightarrow$	ATP	+	MQ ,
R17r_sdh	SUCC	+	0.3 MQ						$\rightarrow$	0.3 MQH_2	+	FUM.

The information about the extracellular protons is thereby stored in the MQ species. As 3 H_e^+ are required to synthesize one ATP and also 3 H_e^+ can be pumped per MQH_2 their number can be directly related. Additionally, for each H_e^+ produced or consumed by one of the reactions a fraction of 0.3 additional MQH_2 is produced.

Simulation and optimization generals

The model was implemented in the D2D framework, whereby the required data and model files were written automatically by additional Matlab scripts to avoid introduction of typing errors. Parameters were initialized within the default bounds as summarized in Table 4.2. For fitting routines, the parameter space was searched logarithmically to improve convergence of the fitting routines. The fitting routine for the detection of regulatory points as described below was implemented in Matlab employing the `fmincon` optimizer. Concentrations were simulated in logarithmic units. Mean active enzyme concentrations, as calculated from the proteomics measurements, were included directly in the kinetic rate laws as fixed values.

6.2 Minimum regulation in steady state

As mentioned before, cells can flexibly reroute their metabolic fluxes in changing conditions. For mycobacteria, this flexibility is also a major pathological treat, allowing them to thrive in a vast variety of host environments. How this redistribution is orchestrated on the mechanistic level has been subject of many studies in various species, recently boosted by omics approaches. In the following Section, a novel model and data driven approach is presented to identify a minimum set of regulatory interventions to transition from one metabolic condition to another. Thereby, the generated data is exploited to calibrate the developed CCMtb model to different metabolic steady states. An optimization strategy is then used to find the least expensive transition between those steady states, measured by the number of required regulatory actions.

6.2.1 How to detect regulatory interactions in metabolism

Previous methods

The quest to detect regulatory edges or likely regulated reactions in metabolic systems has previously been undertaken by a number of experimental and theoretical approaches, which are briefly summarized in the following Section.

Metabolic control analysis [70, 83] and ***supply-demand analysis*** [161] define regulatory points based on an already existing, fully parameterized model. This fact is also the main drawback of these methods, as such models are difficult to obtain in practice and available only for a small set of well studied pathways. In the two methods, regulations are likely if control coefficients are large or a significant change in the supply-demand characteristics of a metabolite occurs in a pathway.

A method branching from metabolic control analysis is (time dependent) ***hierarchical regulation analysis***, dissecting between enzyme expression based hierarchical and substrate dependent metabolic regulation [166, 189, 199]. Intended to complement metabolic control analysis to changes in enzyme expression levels, the analysis is based on the decomposition of the reaction's kinetic law for the flux v into an enzyme dependent term f and a substrate dependent term g: $v = f(E) \cdot g(\mathbf{x}, \boldsymbol{\theta})$ (see also Equation 2.4). By logarithmizing and normalizing to the total flux $\bar{\mathbf{v}}$ of a linear pathway the hierarchical regulation coefficient ρ_h and the metabolic regulation coefficient ρ_m in response to a local perturbation of the enzyme concentration are defined as:

$$\begin{aligned} \rho_h &= \frac{\Delta \ln f(E)}{\Delta \ln \bar{\mathbf{v}}} \\ \rho_m &= \frac{\Delta \ln g(\mathbf{x}, \boldsymbol{\theta})}{\Delta \ln \bar{\mathbf{v}}} \end{aligned} \tag{6.1}$$

Depending on the magnitude of each coefficient, the regulation of a reaction can be classified. The analysis is again based on a detailed dynamic description of the reactions in question.

Chubukov *et al.* [36] experimentally tested the theory of the above regulatory coefficients across a number of culture conditions representing large environmental perturbations based on ***flux measurements*** in *B. subtilis*. They correlated the ^{13}-C labeling derived fluxes with enzyme concentrations estimated from transcript levels to calculate ρ_h. The study concluded that for most of the central carbon metabolism, gene expression changes were insufficient to explain flux changes. While enzymes in the Krebs cycle partly showed a correlation, split reactions at metabolic branch-points were also not significantly expression regulated. Only enzymes that were specifically switched on in single conditions ('On/off state enzymes') showed strong transcriptional regulation. An estimate of the effects of changing metabolite concentrations on

the changes in flux was also calculated as the combined coefficient ρ_{es}, but was small for most considered reactions.

Thermodynamic considerations are also frequently used to propose regulated reactions [71, 98]. The authors argue that the flux through reactions close to equilibrium is sensitive to changes in metabolite concentrations and that these reactions are hence unlikely targets of regulation.

Regulations cannot only be facilitated by changes in enzyme or educt concentrations, but also by modifying the enzyme ***posttranslationally*** to change its catalytic parameters. Current omics techniques can for example measure phosphorylations of enzymes in different growth conditions, a post translational modification (PTM) that is often found to have regulatory effects.

For mycobacteria, available phosphoproteomics data are summarized in Appendix A.23, data on other PTMs, such as acetylations, methylations and others was not available. As PTM-sites of proteins often share steric characteristics in specific domains, a domain search for potentially posttranslationally modified enzymes in the Pfam database [55] was conducted, but did not give any significant results in *M. tuberculosis*. Finally, known regulations of the orthologs of CCMtb enzymes in other mycobacterial species were collected from the BRENDA data base [175] and are summarized in Appendix A.24.

A new data and model driven approach

In the following, a method to predict regulated reactions as a minimum set of impact points to transition from one metabolic phase to another, will be described. The approach combines the experimental proteomics and metabolomics data with the powers of kinetic modeling. Thereby, the contribution of enzyme kinetics to the redistribution will be addressed specifically, neglecting the contribution of thermodynamics for the time being. This is done for reasons of simplification, as thermodynamic-kinetic reaction equations are computationally expensive in the optimization.

The regulation detection method described here is based on the assumption that all flux controlling factors, enzyme concentrations, saturation state, thermodynamic gradients and regulatory actions (see Equation 1.1), will be constant in an established steady state. It can hence be assumed that the effects not included in a mathematical model, such as the thermodynamic and regulatory terms for the CCMtb model here, take a fixed, potentially unknown value in this steady state. Kinetic parameters for each reaction estimated under this assumption will then circumscribe all their contributions. For another nutritional condition, the found parameters will probably fail to describe the data, as the regulatory terms may have changed and are not accounted for. Assuming that the cell, in order to save resources, optimizes to execute as few as possible actions to reroute the flux while transitioning from one steady state to another, an optimization problem can be formulated. Kinetic parameters are optimized to fit steady state metabolite concentrations for several conditions, with regulation parameters introduced for reactions where this is impossible. The number of regulation parameters is simultaneously minimized, to account for optimality in terms of an efficient transition. Thereby, this work focuses on the transition from the metabolic state on one experimental carbon source to another. The different growth phases are neglected to simplify the approach (*i.e.* all glucose data were merged), but could be represented in the same way in an extended study.

6.2.2 Mathematical formulation

Objective function

The best solution to the above described problem has to fulfill three objectives: (i) The model has to reach a steady state, where all time derivatives of the metabolite concentrations $\mathbf{x}$ vanish. This condition was realized by calculating the derivative at of $\mathbf{x}$ with regard to time and including

the metabolite concentrations $\mathbf{x}$ as variables to the optimization. Via this derivative term, the kinetic law enters the objective function. All enzymatic reactions are formulated as convenience kinetics [104], as stated before. (ii) At the same time, the model observables $\bar{\mathbf{y}} = f(\bar{\mathbf{x}})$ have to be close to the measured data in steady state $\bar{\hat{\mathbf{y}}}$. (iii) The regulation is implemented in a very simplified way, by multiplying scaling parameters $\boldsymbol{\theta}^s$ to each kinetic law in each condition as:

$$v_{j,i} = f(E_{j,i}, \bar{\mathbf{x}}, \boldsymbol{\theta}) \cdot \theta^s_{j,i}, \tag{6.2}$$

with i and j as indices for the condition and the reaction, respectively. The objective function punishes deviations from 1 for each of those parameters. Consequently, these parameters will remain at a value of 1 unless their value is required to change to achieve a steady state or agreement with the data. The objective function then reads as:

$$\min_{\boldsymbol{\theta},\boldsymbol{\theta}^s,\mathbf{x}} \sum_{i=1}^{n_c} \left(\underbrace{w^s \cdot \sum_{k=1}^{n_x} \left(\frac{\partial \bar{x}_{i,k}}{\partial t} \Big/ \bar{x}_{i,k} \right)^2}_{\text{(i) deviation from steady state}} + \underbrace{w^y \cdot \sum_{l=1}^{n_o} \left(\frac{y_{i,l} - \bar{\hat{y}}_{i,l}}{\overline{sd}_{i,l}} \right)^2}_{\text{(ii) deviation from data}} + \underbrace{\sum_{j=1}^{2n_r} \log_{10}(\theta^s_{i,j})}_{\text{(iii) minimum regulation}} \right),$$

$$\text{subject to} \quad \boldsymbol{\theta}_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{ub} \text{ and} \quad \mathbf{x}_{lb} \leq \mathbf{x} \leq \mathbf{x}_{ub} \tag{6.3}$$

The parameter vector $\boldsymbol{\theta}$ contains catalytic activities, Michaelis-Menten constants as well as uptake and growth rates for each condition. The indices n_r, n_x, n_o and n_c are the identifiers for enzymatic reactions, metabolites, observables per condition and conditions, respectively. There are a total of $2n_r$ scaling parameters to be optimized, one for the forward and the backward direction of each enzymatic reaction. The boundaries $\boldsymbol{\theta}/\mathbf{x}_{ub/lb}$ restrict possible values for parameters and metabolite concentrations (see Table 4.2). The glucose condition was defined as a reference condition i_{ref} for which all $\theta_{i_{ref},j}$ were fixed at 1. w^s and w^y are the weights defining the contribution of the steady state and the data term to the overall optimality. Weights were adjusted manually to achieve a good convergence behavior of the optimizer. The deviations from the steady state (i) and the data (ii) are squared sums normalized to the calculated standard deviation of each observable in the steady state $\overline{sd}_{i,l}$ and the current values of the metabolite concentrations $\bar{x}_{i,k}$. This normalization avoids biases due to vastly different metabolite concentration ranges.

Technical information

The optimization was conducted outside the D2D framework, as no dynamic simulation is required for steady state estimation. Single functions to calculate the flux vector $\mathbf{v}$ were automatically extracted from D2D to still benefit from the strong formalization and model organization properties of the tool. The optimization procedure included a large number of fits with Latin hypercube sampled initial parameter sets ($\sim$500) and simulated annealing-like repetitions [91]. Specifically, for each initial point several subsequent fits were run ($\sim$50), with the initial point being the found optimum parameter set of the previous fit with some noise added. If no steady state was found, the bounds on single parameters were extended, based on the distribution of found parameter values of all fits. This procedure was repeated iteratively, until a satisfying steady state was found (mean relative deviation from steady state smaller than $5 \cdot 10^{-3}$).

Statistical evaluation of found regulation parameters

To identify the reactions that required significant regulation, the distribution of estimated logarithmic values of each scaling parameter was tested against the Null-Hypothesis that their mean is 0 (*i.e.* the reaction is not regulated) with a t-test. A p-value cutoff of 10^{-5} was chosen. In addition, also the extent of the regulation had to exceed a certain value, such that the resulting mean fold change of the flux (MFC) would be larger than 1.6 (corresponding to $|\log_2(\text{MFC})| > 0.9$).

6.2.3 Results: A purely kinetic model is not enough

As a pre-calculation, the steady state of each nutritional condition – growth on glucose, propionate and glutamate – was fitted individually. Such a setup has the benefits of smaller computational costs, as well as of providing a general feeling for the system and the parameter space. Also, infeasibilities in a single condition can be detected easily. For this purpose, the objective function in Equation 6.3 can be simplified as no regulatory term (iii) is required and the sum over all n_c conditions can be removed.

The three steady states could be fitted by iterative adjustment of general parameter boundaries as described above. For the adjustment, histograms of all found parameter values for a fit series that did not arrive at a feasible steady state can be used: Whenever all fitted values for one parameter were close to a boundary, this boundary has to be changed appropriately. Such boundaries act as limiting constraints in analogy to the shadow prices of the linear optimization problems.

The histograms for the final individual fits that resulted in a feasible steady state while being in good accordance with the data can be found in Appendix A.25. From there it also becomes apparent that for some parameters the optimum distributions do not overlap, *i.e.* that the parameters have to take different values in different conditions. In such cases, regulation is required to transit from one condition to another. Besides the histograms, it is often useful to examine parameter correlations to identify dependencies between parameters. The examples in Figure 6.3 show different types of more or less well defined dependencies. Figure 6.3 D highlights that these can also vary between the conditions, again hinting towards a requirement for regulation.

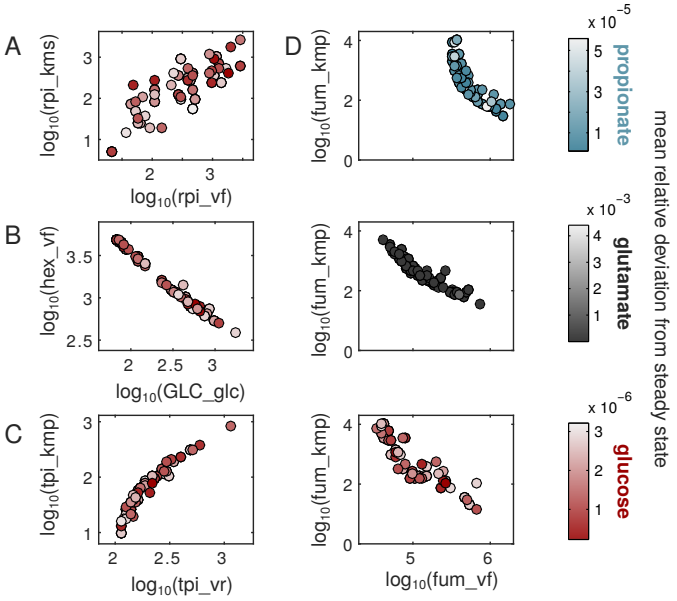


Figure 6.3: Examples for observed parameter correlations in the individual steady state fits. Each dot corresponds to one estimated parameter set, shown are only the 100 best fits. The color coding is according to the deviation from steady state (term (i) in Equation 6.3). Red dots correspond to glucose condition, blue to propionate and gray to glutamate. Each plot shows the estimated values for a pair of parameters. Please note that all axes are in logarithmic scale. **A** weak dependency, **B** strong linear dependency, **C** strong nonlinear dependency, **D** dependencies for one pair of parameters of the fumarase reaction compared between the three nutritional conditions.

For consistency it was tested if the model with purely the kinetic reaction description and without any regulation could reproduce the data of all conditions' steady states. Most unsurprisingly, this was not possible, consistent with the non-overlapping optimal ranges for some parameters in different conditions (see again Figure 6.3 D).

6.2.4 Results: The most probable set of regulations

Likely regulated reactions

Starting from the parameter sets found in the individual steady state fits, the minimum regulation fit was conducted. Regulation parameters were initialized with 1 (0 in logarithmic scale). The final estimated regulation parameters $\boldsymbol{\theta}^s$ are summarized in Figure 6.4 A and mapped to the CCMtb network in Figure 6.4 B. Most $\boldsymbol{\theta}^s$ remained at their initial value and did not contribute to overall regulation. Significant regulations were found in upper glycolysis and the Krebs cycle.

Biological recovery of known regulatory points in Mtb

Compared to known regulatory points in *M. tuberculosis* (Figure 6.4 C) the approach could recover only a subset of regulations. Part of the Krebs cycle (fumarase, malate dehydrogenase, citrate synthase, isocitrate dehydrogenase and alpha-ketoglutarate decarboxylase) as well as glycolytic reactions (fructose-1,6 bisphosphate (FDP) aldolase, triosephosphate isomerase and pyruvate kinase) are contained in both sets. Especially regulation in the pentose phosphate pathway and the anaplerotic reactions could not be detected here.
Explanations for this discrepancy could be that most of the known regulations were measured in conditions very different from the experimental ones in the datasets used here. Inhibitions and activations are mostly characterized by *in vitro* enzyme assays and might just be inactive in the conditions considered here. The same holds for the phosphorylations, that are additionally only candidates for regulation – their real function is often unknown.
Although there are some discrepancies, some common features of regulation occur in both the known and the here predicted set of regulations. The PEP-pyruvate-oxaloacetate node, for example, has been characterized as important flux control point in bacteria [172] and was also found to be regulated in mycobacteria in various ways (see Figure 6.4 C). Here, its Krebs-cycle part is found to be regulated along with the glycolytic pyruvate kinase reaction.
The gluconeogenetic fructosebisphosphatase glpX is inhibited by methylcitrate [51]. Despite the vastly different concentrations of this inhibitor measured by metabolomics here, the reaction is not classified as regulated in propionate conditions. However, both adjacent reactions (FDP aldolase and phosphoglucose isomerase) are found to be regulated. This could hint to identifiability issues of the optimization, where the specific reaction in this linear stretch cannot be detected, but the overall need for regulation is conserved. As a last example, FDP has been characterized as a global sensor for glycolytic flux in bacteria [35, 96]. Also here its producing and consuming reactions are consistently regulated between glycolytic and gluconeogenetic conditions.

6.2.5 Discussion

An automatized method to predict regulatory impact points

The presented approach combines high resolution metabolite and proteomic data with mathematical modeling and a cellular optimality concept to detect regulatory impact points. In contrast to other experimentally based methods, no expensive flux measurements are required. Also no model fully covering the dynamic behavior is needed as in previous approaches, but rather

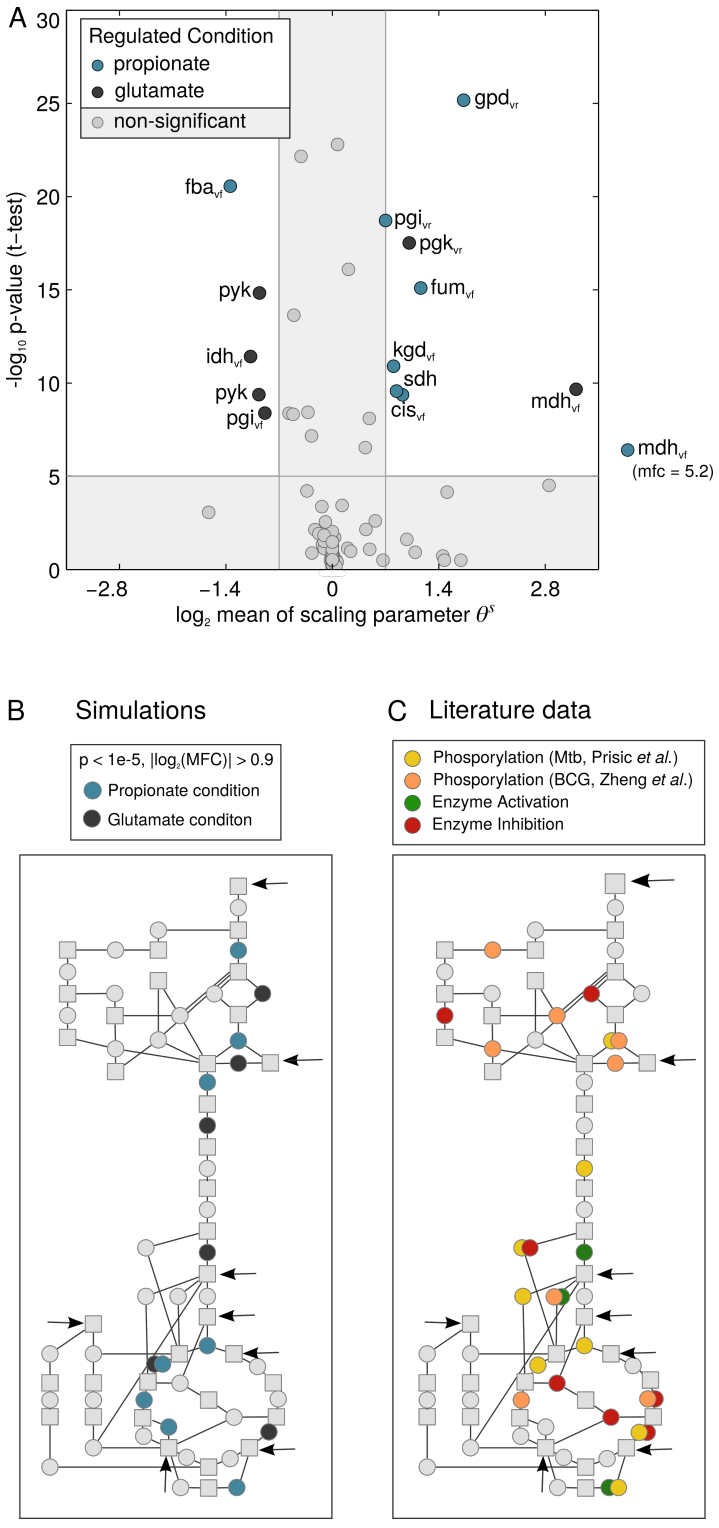


Figure 6.4: Minimum set of regulatory interventions as found by the optimization. **A** Volcano plot of the regulation parameters $\boldsymbol{\theta}^s$, plotted is the mean of the distribution of all found values against the p-value obtained as described in the text. Parameters in the gray area are not significantly regulating their corresponding reactions. Significant parameters are labeled with the name of the reaction and its direction (vf - forward, vr - backward). **B** Significant regulations mapped to the network, color code as above. As before, squares denote metabolites, circles denote reactions. **C** Known regulations from literature mining, a detailed list can be found in Appendix A.23.

a range of feasible parameterizations is explored. Thereby, the reduction to the steady state drastically reduces the computational effort – at the cost of neglecting dynamic information on regulations. The method can also not dissect thermodynamic or regulatory actions, which will both affect the reaction flux in addition to the kinetics. Enzyme expression changes are directly included via proteomics measurements and do hence not confound the results.
The approach is easily extendable to other metabolic systems, organisms or additional data sets, as it is highly formalized and therefore easy to reproduce. Due to the employment of convenient, scalable kinetic laws, no prior knowledge on the enzyme kinetics was required.
The fact that not even the steady state of the central carbon metabolism, as a very reduced system, can be described by a purely kinetic model, highlights the necessity of knowing and including regulatory actions in kinetic models. Regulation can there, as well as probably in the cell, also be seen as a way to overcome strict substrate dependencies of the fluxes and allow for a larger spectrum of possible flux distributions. The results presented here allow for their detection in an integrative and scalable way. The found regulatory points can now be tested experimentally as a proof of their functionality or existence. Such a targeted approach requires much less experimental effort than for example metabolome wide screenings.

The steady state results constrain the parameter space

Computationally, the estimation of parameters for a system in steady state is much less expensive than calibrating the model's dynamics. For each optimizer step, the derivative function (Equation 2.3) has to be evaluated only once, while in the dynamic case, its whole time evolution has to be simulated numerically. However, as the steady state problem is much less constrained than the dynamic one, it usually has many feasible solutions, which contain less information on the system. In other words, many of the parameters will be non-identifiable and directly dependent on each other. Nevertheless, the fact that the solutions to the dynamic problem have to also comply with the steady state solutions can be exploited to constrain the feasible parameter region of the dynamic problem. The simplest way to do so would be to use the estimated steady state parameter sets as initial points for the dynamic parameter estimation, alongside a formulation for the steady state constraint at $t = 0$ of the dynamic simulation (before the nutrient shift occurs). A more sophisticated way would be to reformulate the model equations, such that each dependent parameter is expressed as an algebraic equation of the corresponding independent parameter. However, this requires the correct mathematical description of the dependency, for example an equation for the apparent curves in Figure 6.3, which is a non-trivial problem.

Possible extensions and improvements

The approach as shown here can be easily extended to cover transitions between distinct environmental conditions, not limited to switching nutrients. For example with the data acquired in this work, also the transition between different growth phases could be investigated.
The method could also be extended to directly include thermodynamic effects via a suitable rate law, which would allow to separate thermodynamic from regulatory effects directly. Furthermore, mechanistic accuracy can be improved by not only scaling the entire forward or backward flux, but also by scaling of the Michaelis-Menten constants. Especially allosteric regulation and competitive inhibitions change the substrate affinities and hence the saturation behavior.
A further evolutionary aspect could be to link regulatory parameters between the conditions. If the cell already implemented the regulation in one condition, it is more likely to appear in another, as probably not every enzyme has evolved to be regulated.
Technically, the detection of significant parameters could be better assured by profile likeli-

hoods, relieving the bias of a potentially insufficient number of fits conducted. However, their calculation is computationally expensive, as an additional optimization has to be carried out for each point of the profile of each parameter. To resolve the bias introduced by the weights of the single components of the objective function, the steady state and data compliance could also be implemented as constraints on the optimization problem, resulting in only one objective of minimizing regulation. However, the information of the distribution of the measured metabolite concentrations would then be lost.

7. Pdtx: Propionate detoxification in mycobacteria

Which concepts facilitate metabolic flexibility in mycobacteria?

While the previous Chapters focused on metabolic steady states and what can be learned from them, metabolite dynamics are considered in the following Chapter. Dynamic data usually contains more information on a system in question, but is also more difficult to attain. Here, the dynamic metabolite data from the shift experiments will be used to characterize the individual treats of two seemingly redundant metabolic pathways relevant for the pathogenicity of *M. tuberculosis*. The findings highlight the capabilities of integrated kinetic modeling in contrast to steady state approaches: While based on steady state thermodynamics and measured enzyme concentrations, the distinct functionalities of both pathways cannot be resolved, thermodynamic-kinetic modeling identifies one pathway as a thermodynamically limited but highly responsive rescue pathway while the other is classified as a high-capacity catabolic route.
The presented analyses thereby combine known methods, of which each is adapted and improved for their purpose here, to fully characterize both pathways. The kinetic model integrates dependencies of the flux on the saturation state, the thermodynamic gradient as well as on the measured enzyme abundances (*cf.* Equation 1.1). Especially thermodynamics are neglected in many published kinetic models of metabolism, such that – together with the high data coverage – the model presented here advances the field of metabolic modeling. The described set of analyses can therefore be used as a road-map for the study of similar metabolic systems.

7.1 Propionate metabolism in mycobacteria

Lipids as carbon sources

The major carbon source of intracellular mycobacteria is unknown, but experimental evidence suggests a mixed diet of amino acids and other unknown substrates containing two or three carbon atoms [22]. Such compounds are for example the end products of fatty acid and cholesterol catabolism, which has been characterized as an important intracellular nutrient, especially during dormancy [142, 202]. Pathways required for the catabolism of intracellular carbon sources are promising targets for drug development [22, 138, 202].

Two parallel pathways for propionate catabolism

Propionyl-CoA (PCO) is an exemplary three carbon compound. It can feed carbon and energy metabolism, but is also required to produce relevant mycobacterial lipids and hence has to be synthesized *de novo* if it cannot be scavenged from the carbon source [108]. PCO can be catabolized via two parallel pathways in mycobacteria, the methylcitrate (MCIT) and the methylmalonyl (MMCO) pathway (Figure 7.1). These two seemingly redundant pathways can both fulfill the task of fueling the central metabolism from this non-canonical carbon source. However, in most

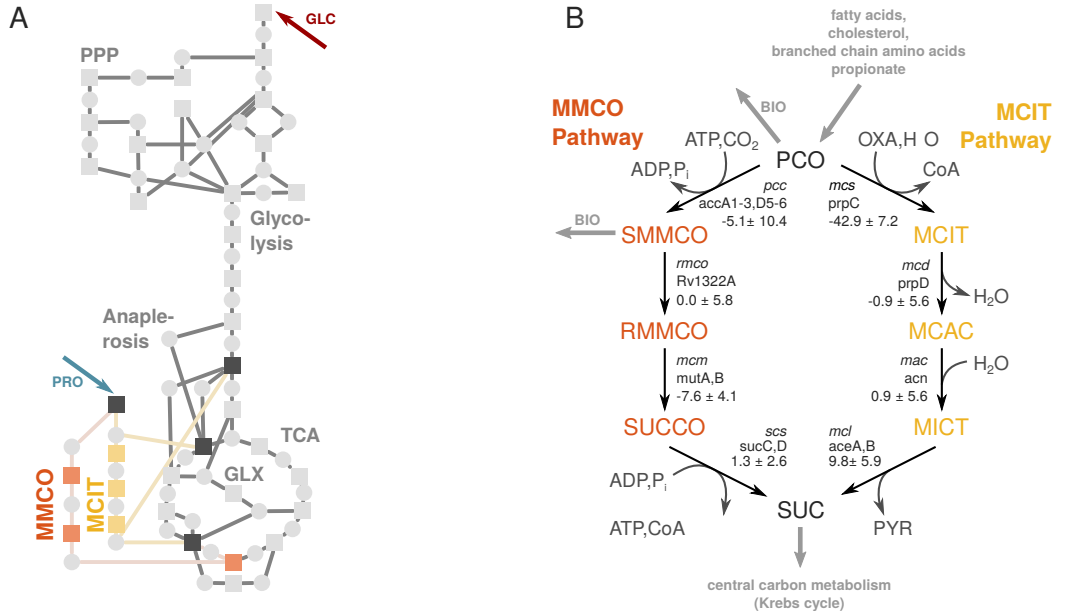


Figure 7.1: Context and structure of the two propionate catabolizing pathways. **A** Location of the two pathways in the central carbon metabolism of *M. tuberculosis*. MMCO pathway intermediates are shown in orange, MCIT intermediates in yellow. Dark gray squares indicate boundary metabolites of the pathways. PPP - pentose phosphate pathway, TCA - Krebs cycle, GLX - glyoxylate shunt **B** Detailed topology of the two pathways. Dynamically modeled species are shown in the respective colors, constant metabolites in gray. For each reaction, the reaction name, the gene name and the standard Gibbs energy $\Delta_r G^0$ are shown. Fluxes to biomass are indicated as gray arrows. For abbreviations see Appendices A.18 and A.29

experimental scenarios requiring this catabolic activity, *M. tuberculosis* prefers the MCIT over the MMCO pathway. Accordingly, the MCIT pathway has been subject of extensive study as an anti-tubercular drug target, encouraged by the lack of a homologue pathway in the human host. Both pathways can maintain flux from PCO to succinate, an intermediate of the Krebs cycle, using different mechanisms. The MMCO pathway depends on ATP activation in the first reaction step (Figure 7.1, *pcc*) allowing for the fixation of one molecule CO_2. The pathway includes one strictly vitamin B_{12} dependent reaction, a complex cofactor that can be synthesized by the bacteria themselves [173, 211]. The ATP is regained in the last step of the pathway, which produces succinate. The pathway also fulfills biosynthetic tasks in the production of surface exposed lipids [80], for which S-methylmalonyl-CoA is used as a building block.
The MCIT pathway, in analogy to the glyoxylate cycle in the central carbon metabolism, invests one oxaloacetate molecule to convert PCO to pyruvate under the release of succinate, which feed back into lower glycolysis and the Krebs cycle, respectively. It was found to be up-regulated in intracellular bacteria *in vivo* (mouse lungs and human macrophages [160, 174, 183, 188]), in dormancy models [177], on *in vivo*-like carbon sources such as cholesterol [68] and is essential in various other experimental contexts [21, 74, 126]. Recently, also its role in antibiotic tolerance by specifically rerouting central carbon metabolic fluxes was described [130]. The MCIT pathway also occurs in other pathogens and is up-regulated upon exposure to propionate, which is often part of their diet in the pathological context (*Salmonella enterica* [140], *Toxoplasma gondii* [107]).

The regulation of cytosolic levels of the intermediates of the MCIT pathway is crucial, as they can become cytotoxic at high concentrations: Two studies [198, 202] showed that inhibition or knock-out of early MCIT enzymes can restore growth of mutants lacking an enzyme for a downstream MCIT reaction (*mcl*) in media containing propionate, arguing for a more prominent effect of the MCIT pathway intermediates' toxicity than the loss of its carbon acquiring function caused by the knock out. Mechanistically, MCIT intermediates at high concentrations inhibit pyruvate dehydrogenase activity [99], which could be the main reason for their toxicity. While all other reactions of the two pathways are catalyzed by a dedicated enzyme, the two final reactions of the MCIT pathway are carried out by central metabolic enzymes: isocitrate lyase (aceA/B, [67]) from the glyoxylate shunt and aconitase (acn) from the Krebs cycle, highlighting the tight interconnection of the MCIT pathway and the control of its intermediate levels with the central carbon metabolism.

The interplay between the two pathways

If the predominant MCIT pathway is lost, as for example in knock-out experiments, the MMCO pathway can partly compensate for the loss [51, 173]. This compensation can, however, only be achieved if its flux is increased, for example by the addition of vitamin B_{12}, an essential cofactor of the *mcm* reaction. Similarly, vitamin B_{12} can boost the growth of the wild type on propionate, by facilitating an additional carbon feed. However, even with the addition of vitamin B_{12}, growth limiting intracellular PCO concentrations are reached if more propionate is added to the medium [51], arguing for a upper capacity limit of the MMCO pathway. This compensating behavior could explain why the deletion of the first two enzymes of the MCIT pathway abrogates the ability to grow on propionate *in vitro*, but not in *in vivo* mice studies [128], where the cells are likely exposed to lower PCO concentrations, which can still be catabolized via the MMCO pathway.

As mentioned before, the methylmalonyl-CoA pool controls the synthesis of virulence relevant surface exposed lipids, such as phthiocerol dimycocerosates (PDIMs) and Sulfolipid-1 . Hence, the MMCO pathway could possibly "buffer away" the excess PCO in cells with a dysfunctional MCIT pathway towards the production of those lipids. Accordingly, also other genes of the draining pathways into these lipids were found to be essential in media requiring PCO detoxification [80]. As the diverted carbon could then not be used in the central metabolism, utilization of this metabolic route would contribute less to the bacterial growth. As a limitation, also fatty acid precursors derived from acetyl-CoA are required for this detoxification [99].

7.2 A model for robust detoxification via two parallel pathways

In the following the question why mycobacteria keep two seemingly redundant pathways for the same task of propionate catabolism is addressed by means of thermodynamic and kinetic modeling on different levels of resolution, integrating high quality proteomics and metabolomics data. Unique traits of each pathway are characterized to explain why both are required and why the bacteria favor the MCIT pathway despite the danger of self-intoxication by its intermediates. The flux through a reaction or pathway is shaped by the enzyme expression but also by metabolite dependent enzyme saturation and reaction thermodynamics and regulation (see Equation 1.1 and other works [46, 48, 57, 92, 134]). Sub-optimality in any of those factors in the cellular context can render a pathway unfavorable.

Here, established modeling approaches are used to test the significance of each of them for the two pathways in question. Firstly, pathway thermodynamics [134] are employed to show

that the thermodynamic gradient alone is insufficient to explain the preference of the MCIT over the MMCO pathway in intracellular conditions. Also, the subsequent inclusion of specific enzyme masses, with heavier enzymes signifying a higher cellular cost per unit of flux, does not characterize the MCIT pathway as preferable. Therefore, a full dynamic description was developed based on the nutrient switch experiments and thermodynamic-kinetic modeling. The model includes enzyme expression, thermodynamics and saturation and allows to identify the specific traits of each pathway: While the MMCO pathway can quickly react to sudden changes in the propionate concentration and detoxify the cell accordingly, the MCIT pathway can reach a higher overall flux capacity.

7.2.1 Results: Pathway thermodynamics and enzyme costs dissected

The MMCO pathway has a higher thermodynamic driving force and can run reversely

First, to test whether the MCIT pathway might be preferred *in vivo* due to a higher thermodynamic driving force, the thermodynamic gradients within the two pathways were analyzed. The maximum-minimum driving force, as introduced in Chapter 5.2, (MDF, [134]) describes the maximum of the thermodynamic driving force -$\Delta_r G'$ of the least favorable reaction within a metabolic pathway. It is used to identify thermodynamic bottlenecks within a pathway context, based on optimizing metabolite concentrations within sensitive, physiological bounds (see Table 4.2). Extending the original method, uncertainties of the Standard Gibbs Energies $\Delta_r G^0$ as

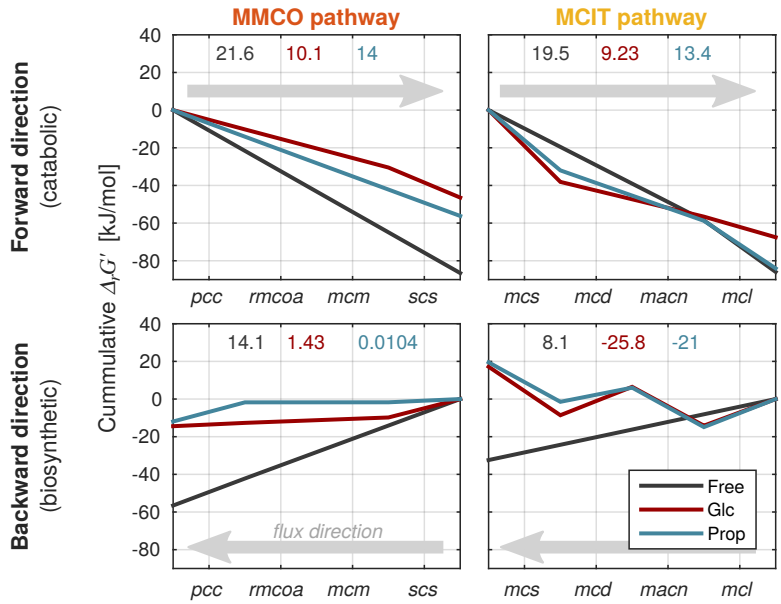


Figure 7.2: Driving force profiles after the maximum-minimum driving force optimization. The cumulative Gibbs energy $\Delta_r G'$ along the pathway is plotted, the MMCO (left) and the MCIT (right) pathways are analyzed in forward (upper panel) and backward (lower panel) direction as indicated by gray arrows. The participating enzymes are labeled on the x-axis. Monotonously decreasing lines in the direction of flux imply pathway feasibility, increasing lines mark infeasible reactions. The different sets of bounds on the metabolite concentrations are shown in different colors: gray - default physiological bounds ('Free'), red - measured concentrations under glucose conditions ('Glc'), blue - measured concentration under propionate conditions ('Prop'). Color coded numbers in the top of the plots indicate the corresponding MDF value (*i.e.* the driving forces of the least favorable reaction), negative values indicate pathway infeasibility.

well as the concept of observables were included here, resulting in a nonlinear optimization problem (Appendix A.26).
The MDF for each of the two pathways was calculated under three different conditions characterized by a set of specific metabolite concentrations: First unbiased metabolite concentration bounds in the default range of 1 μM - 50 mM (see Table 4.2) were applied. To capture specific traits of catabolic and biosynthetic conditions, measured steady state metabolite concentrations of the propionate or glucose cultures were included as two additional sets of environment specific concentration bounds.
Both pathways show comparable MDFs in forward (catabolic, propionate consumption) direction but only the MMCO is feasible in the backward (biosynthetic, glucose consumption) direction under physiological conditions (Figure 7.2, red and blue lines). The prohibiting reaction of the reverse MCIT pathway is the final splitting of methylcitrate into PCO and oxaloacetate by the *mcs* reaction. Notably, only the measured metabolite concentrations reveal the infeasibility, under default concentration bounds the MCIT could run reversely (Figure 7.2, lower right, gray line).
Its lower MDF in the catabolic direction would argue for the predominant usage of the MMCO pathway, contradicting the previously described preference for the MCIT pathway. However, the MDF differences are not very prominent, especially given the uncertainties in metabolomics and $\Delta_r G^0$ measurements.

Protein costs are condition dependent and favor different pathways in different nutritional conditions.

As the thermodynamic gradient alone cannot explain the pathway preferences, the enzyme mass was included additionally in the optimization, to asses if higher protein costs due to large enzymes could be limiting the MMCO pathway. Protein cost optimization has been studied in many contexts before [48, 57, 92], here the method of Flamholz *et al.* [57] was adapted, dissecting the effects of enzyme expression, saturation and thermodynamics. The optimization is based on a decomposition of the Michaelis Menten rate law for the flux v into multiplicative contributions of the enzyme amount E, the catalytic activity k^{cat}, a purely educt dependent saturation term and an educt and $\Delta_r G^0$ dependent thermodynamic term that quantifies the extent of futile backward flux ([136], *cf.* Equation 1.1).

$$v = \underbrace{E}_{\text{enzyme}} \cdot \underbrace{k_+^{cat}}_{\text{catalytic activity}} \cdot \underbrace{\left(\frac{\prod_i (S_i/k_{s,i}^M)}{\prod_i (S_i/k_{s,i}^M) + \prod_j (P_j/k_{p,j}^M) + 1} \right)}_{\text{saturation term}} \cdot \underbrace{(1 - e^{\Delta_r G'/RT})}_{\text{thermodynamic term}} \quad (7.1)$$

Based on this decomposition, optimal protein costs can be estimated as the minimum possible enzyme investment per unit of flux. This cost will increase if the enzyme is under-saturated or has a shallow thermodynamic gradient, both depending on the metabolite concentrations. The original method as applied by Flamholz *et al.* did not account for different enzyme masses, which were included here specifically as well as the contribution of product saturation to the kinetic term (for the mathematical formulation of the optimization problem see Appendix A.27). As the MCIT pathway was characterized as infeasible in the biosynthetic direction in the MDF optimization, only the catabolic, propionate consuming direction was analyzed here.

The enzyme costs are comparable between the pathways (Figure 7.3 A) in physiological conditions. The MMCO pathway requires slightly less protein under glucose conditions, whereas the MCIT pathway would be preferable under propionate conditions. For all reactions, the

contributions to the total protein costs arising from the cost of under-saturation (green bars) are higher than from the cost of backward flux (yellow bars), which would indicate a kinetic limitation rather than a thermodynamic one. The saturation term is, however, largely biased by the assumption of uniform kinetics for all the enzymes involved (all k^{cat}= 100 s^{-1}, all k^{M}= 100 μM, [57]), as there are no measured kinetic parameters for the pathways' enzymes (with the exception of *pcc* and *mcl*). Consistently, the calculated minimal cost enzyme abundances λ do not correlate with protein abundances measured in the two culture conditions (Figure 7.3 B).

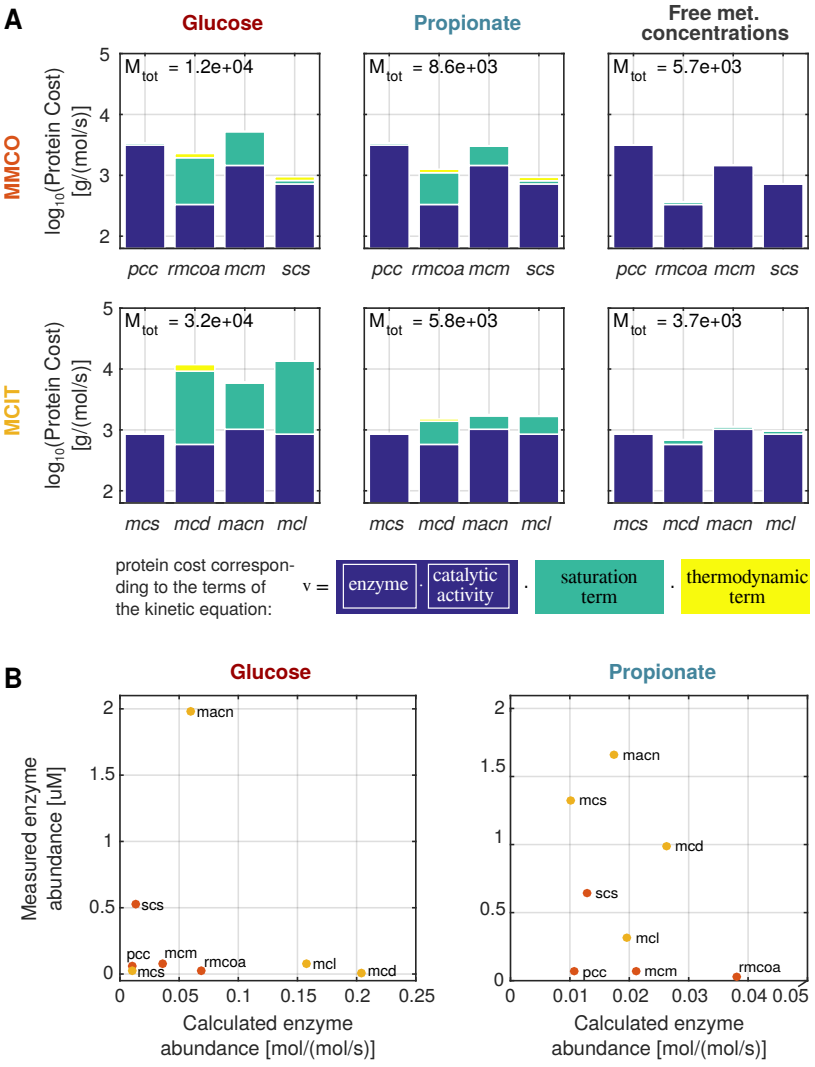


Figure 7.3: Theoretical pathway enzyme costs. **A** Minimum enzymes cost for MMCO (upper panel) and MCIT (lower panel) pathways under different growth conditions (*i.e.* sets of metabolite concentrations, columns), dissected in the contribution of the baseline (dark blue), under-saturation (turquoise) and backward-flux (yellow) costs. **B** Predicted enzyme amounts per unit or flux compared to the measured enzyme concentrations in 2 growth conditions. The measured value represents fully functional complexes with one active site (*cf.* Appendix A.30).

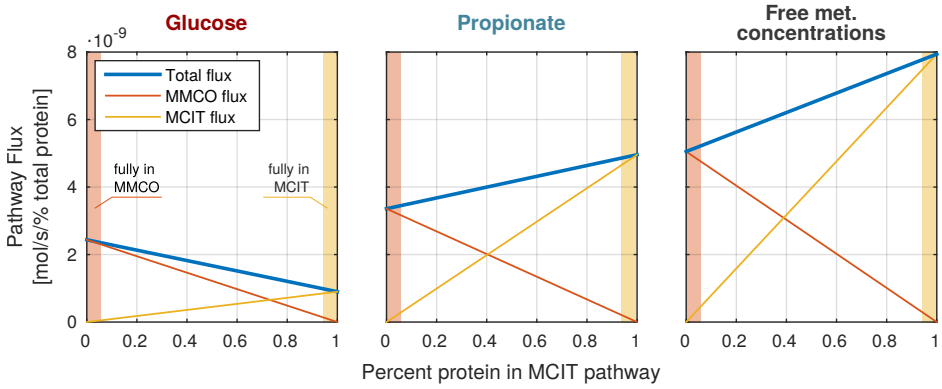


Figure 7.4: Maximum pathway flux for limited enzyme availability. We allowed 1 % of the total *M. bovis* BCG protein mass to be distributed between the two pathways in different ratios (x-axis, relative to the MCIT pathway). Hence, the bounds of each plot represent the full amount of protein being contained in one of the pathways (highlighted in the respective colors). The plots show the contribution of the MMCO (orange line) and the MCIT (yellow line) to the total flux (blue line) for each ratio (11 sampling points, *cf.* Figure A.28) for the metabolic conditions used before: Glucose (left), propionate (middle), default physiological concentration bounds (right).

Flux maximization with limited enzyme availability

An approach complementary to the enzyme cost calculation would be to maximize the flux through a pathway given a limited amount of enzyme available (Appendix A.28). If the constraint on the enzyme amount is applied two both pathways simultaneously, interdependencies in the flux distributions of the two pathways can be detected. For example, a change in a metabolite concentration could slightly lower the flux through one pathway, but dramatically increase the flux through the other due to the differences in $\Delta_r G'$ of the reactions consuming and producing that metabolite. In this scenario, the optimum solution would favor a metabolite concentration that increases the total flux, while accepting a lower flux in one of the pathways – a scenario that would not be detectable when optimizing the protein cost for one pathway alone.
Accordingly, the total flux through both pathways with the three above used metabolic conditions was optimized by scanning a range of possible ratios of enzyme mass distribution between them (Figure 7.4). The same preference for the MMCO pathway under glucose conditions and for the MCIT under propionate conditions can be seen as a result, consolidating the findings from the protein cost optimization (including the same biases with regard to the assumed kinetic parameters). The total flux through both pathways depends linearly on the ratio of total enzyme mass and no interdependency effects could be detected, consistent with their additive behavior in wild type *M. tuberculosis* cultures [51].

7.2.2 Results: Dynamics make the difference

Dynamic detoxification of propionate pulses

The insufficiency of the thermodynamic and enzyme optimality approaches to describe the pathway preferences satisfactorily, suggests that the functional differences might only be detectable in the dynamic response of the two pathways. To test this hypothesis, the dynamic shift data between the two above used conditions were used: From steady state growth on propionate to glucose media, in which reverse flux through the pathway in the biosynthetic direction will be required, and from growth on glucose to propionate media, where a fast detoxification of rising PCO levels will be required.

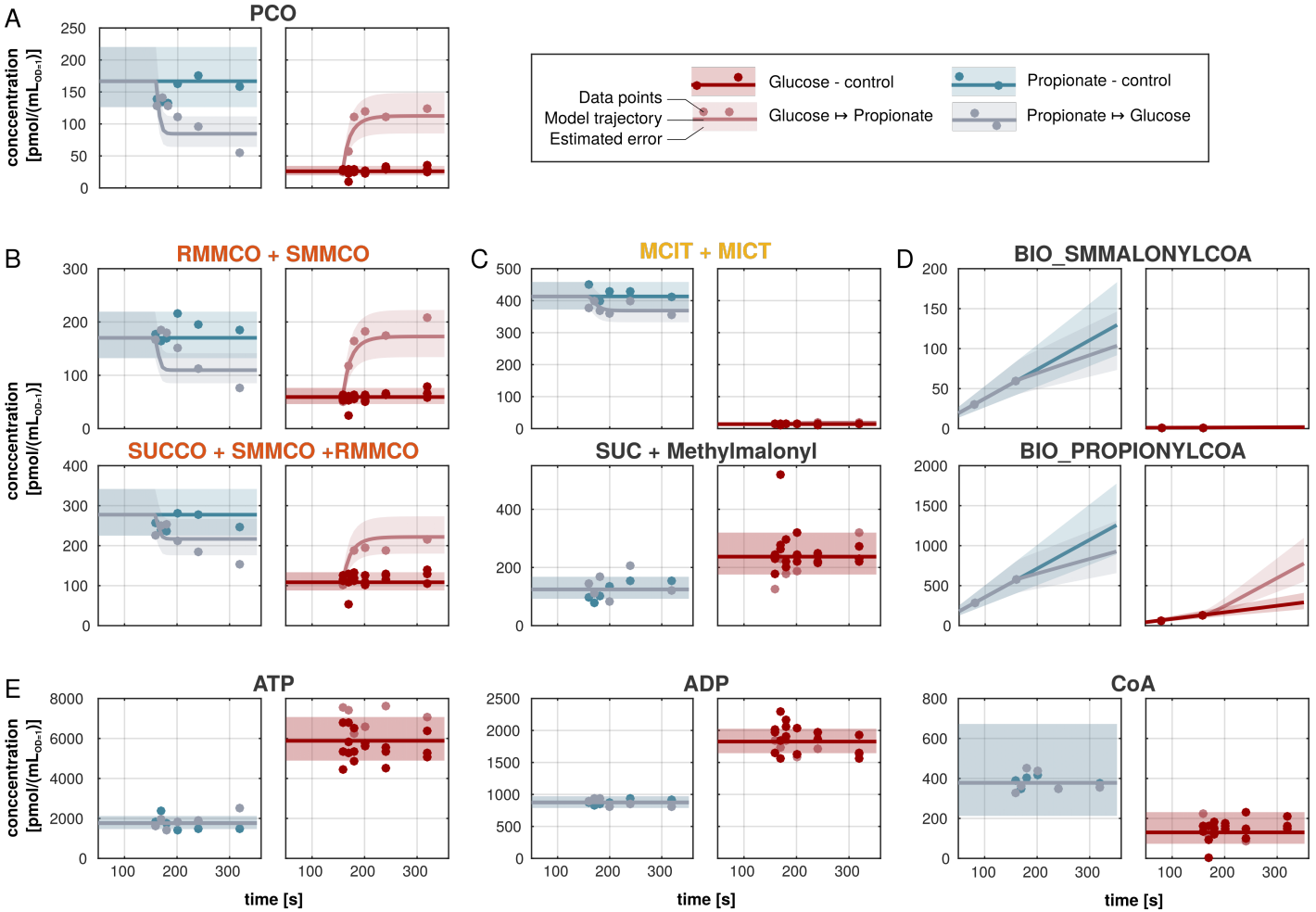


Figure 7.5: The MMCO pathway reacts dynamically to changes in the propionate supply. Experimental metabolomics measurements (dots) along with dynamic model simulations (lines) and the estimated error (shaded areas) for different nutrient switches at time t=160s. **A** Initial species PCO which is the entry point of propionate to the central carbon metabolism. **B** Observables of intermediate species of the MMCO pathway **C** Observables of intermediate species of the MCIT pathway and for the endpoint of the two pathways (succinate, SUC) **D** Dummy biomass species required to guarantee appropriate fluxes draining into biomass production. **E** Co-factors, which are held constant in the model.

From the data (Figure 7.5, dots), it becomes evident that only the MMCO pathway (Figure 7.5 B) reacts dynamically to sudden increases as well as decreases in the PCO concentration (Figure 7.5 A). The MCIT pathway shows only a minor dynamic reaction, but its overall metabolite levels strongly change with the carbon source (Figure 7.5 C).
To explain the observed behavior in detail, a full dynamic model with thermodynamically consistent rate laws was developed (for details see Appendices A.29, A.30). The model was fitted to the dynamic metabolomics data and included also the measured protein concentrations for each nutritional condition. It was able to reproduce the experimental behavior (Figure 7.5, lines and shaded areas as error estimates, corresponding fluxes are shown in Appendix A.31).

Analysis of the calibrated model

Parameter profile likelihoods [97] reveal that most parameters are non identifiable (Appendices A.32, A.33), which is surprising, given the good data coverage. The data situation does hence not allow to pinpoint values for most of the enzymatic parameters. However, general features can be computed from a large number of parameter sets that achieve objective values smaller than the likelihood threshold $\mathbf{LL}^* + \Delta_\alpha$, even if single parameters are not identifiable. The forward catalytic efficiencies, signifying the quotient of k_+^{cat} and k^M for a given reaction substrate pair, are for example well defined for all model reactions (Figure 7.6). Consequently, it is possible to assign a value for them with high confidence, while the individual parameters stay non-identifiable.
The error model parameters $\boldsymbol{\theta}^e$ are usually well defined and easy to estimate. Cases where they are not identifiable within the given bounds, which are commonly based on the experimental error estimates, could hint to a biological behavior which is not accounted for in the model. Here, this was observed for the CoA species, where the estimated error hits its upper bound (see appendix A.32). This discrepancy occurs because this cofactor is held constant in the model, but shows dynamic changes in the glucose measurements (Figure 7.5 E).

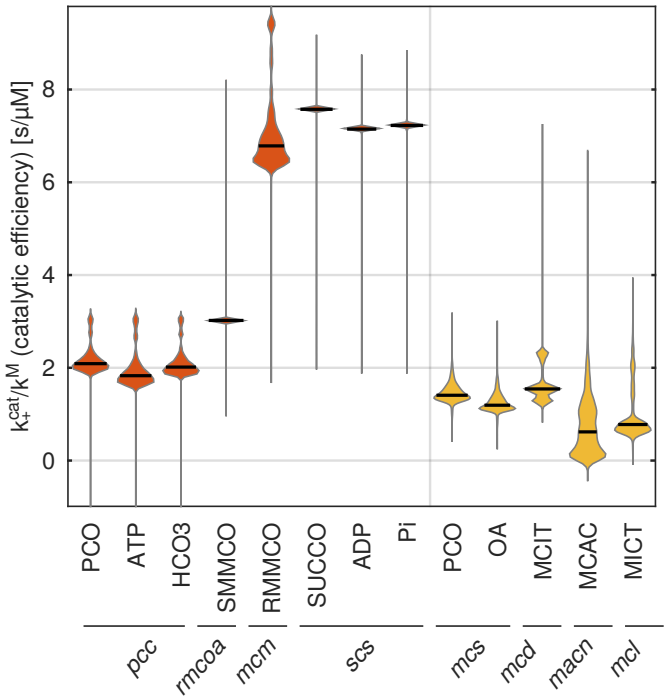


Figure 7.6: Distributions of values for the forward catalytic efficiencies for all enzyme-substrate pairs. Shaded areas describe kernel density estimates for the distributions, the median is indicated by a thick black line. The distribution is calculated from all feasible profile likelihood fits (each corresponding to a black dot in Appendix A.32)

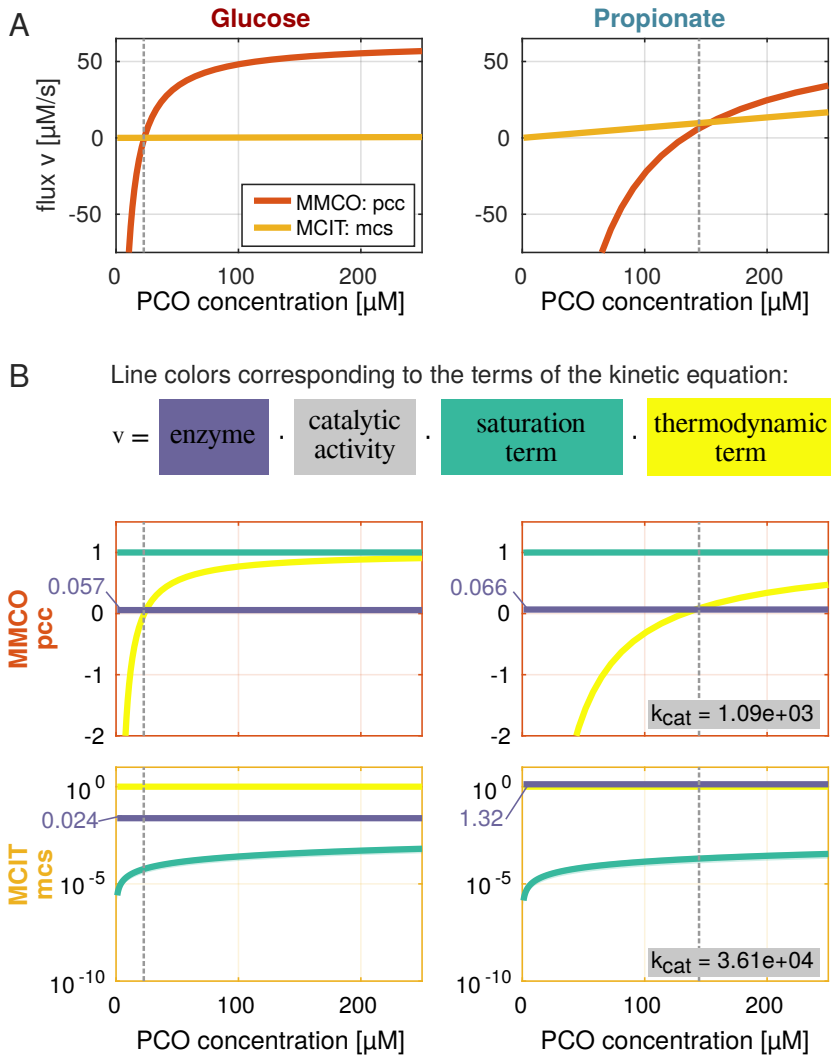


Figure 7.7: Substrate dependency of the first reaction step of both pathways dissected. The two experimental backgrounds glucose (left panel) and propionate (right panel) are shown with the dashed gray line highlighting the mean PCO concentration of the metabolomics measurements. All simulated values are means of all feasible parameter sets of the profile likelihood estimation. **A** Michaelis-Menten-like curves of *pcc* (orange) and *mcs* (yellow), the first reaction steps of the MMCO and MCIT pathways, respectively. The dependency of the flux on the concentration of PCO is plotted. Note that the flux can become negative owing to the reversible nature of the thermodynamic rate law. **B** The total flux as depicted in A, decomposed in the contributions of the individual terms of the reaction rate: Enzyme concentration as measured by proteomics (dark blue, for clarity values are shown on the axes), kinetic term representing the saturation state of the enzyme (green) and the thermodynamic term quantifying the extent of the backward flux (yellow). The thermodynamic and the kinetic term are bounded by 1. The product of the three terms and the k_+^{cat} (lower right corner, right panel) will give the values plotted in A. Upper panel shows dissection for the MMCO reaction *pcc*, the lower for *mcs* from the MCIT pathway. Note that the MCIT terms are plotted on a logarithmic scale.

The decay of MMCO intermediates is not well captured by the model (Figure 7.5 A/B light blue lines). While the data points seem to show a roughly linear decrease in concentration, the model trajectory describes a faster adaptation to the new steady state. A reason for this could be that the step input function does not approximate the propionate depletion well, *i.e.* that the removal of propionate in the medium does not translate directly to zero influx into the PCO pool.

Thermodynamics limit the MMCO pathway flux

The calibrated model allows to analyze the functional flux distribution of the two pathways in detail, to identify the limiting factors for each pathway and condition. To understand how a change in the PCO concentration would impact on the flux in a given metabolic background, the parameterized kinetic equations can be exploited. Here, the Michaelis-Menten curves of the first enzyme of each of the two pathways were examined (Figure 7.7 A), which already highlight the limitations of each pathway. The modular character of the thermodynamic rate law also allows to dissect the contributions of changes in enzyme concentration, enzyme saturation and thermodynamic limitations to the curve (Figure 7.7 B). To avoid a bias due to a local optimum of the parameter estimation, the mean curve of all feasible ($\mathbf{LL} < \mathbf{LL}^* + \Delta_\alpha$) parameter sets found during the calculation of parameter profile likelihoods was calculated (see Appendix A.32).
From the Michaelis-Menten curves it becomes evident that under biosynthetic conditions (glucose consumption, Figure 7.7, left), the MMCO flux (orange) tightly follows the PCO concentration, with the potential to quickly switch from negative, biosynthetic to positive, catabolic rates around the measured PCO concentration (gray dotted line). This switch is solely facilitated by the adaption of the thermodynamic gradient of the reaction (Figure 7.7 B, upper left, yellow line). The flux through the MCIT pathway is almost insensitive to changes in the PCO concentration as the first enzyme is strongly under-saturated at the lower PCO levels (Figure 7.7 B, lower left, green line). Hence, a sudden PCO pulse, as in the experiments, will be solely drained via the MMCO pathway, which rapidly shifts its reaction direction.
Under catabolic conditions (Figure 7.7, right), the MCIT reaction has a higher capacity due to a drastically increased enzyme level (Figure 7.7 B, lower right, dark blue line) and hence out-competes the MMCO reaction. In the catabolic background, the *pcc* reaction has a shallow thermodynamic gradient (Figure 7.7 B, upper right, yellow line), most likely due to the decreased level of the co-substrate ATP. It would even invert its direction upon a slight decrease in the PCO concentration, consequently not fulfilling any catabolic tasks. Here, as well as in the biosynthetic case, the MMCO flux quickly saturates at increasing PCO concentrations in the physiological range, explaining the upper capacity bound for the pathway observed experimentally [51]. The corresponding MCIT enzyme is still in the linear regimen of the Michaelis-Menten curve, guaranteeing a proportional change of the flux with the PCO concentration.
In summary, the thermodynamic properties allow the MMCO pathway to react quickly to changes in the PCO concentration, but also limit its overall capacity, rendering it insufficient to supply enough carbon for bacterial growth. The MCIT pathway, on the other hand, gains momentum only if its enzyme expression is increased, thereby reaching a higher capacity and avoiding saturation effects to efficiently channel carbon to the central carbon metabolism.

Metabolic control is executed mainly by the first enzyme of the MCIT pathway

Employing metabolic control analysis [70] also the question why the bacteria tune the metabolic flux distribution by up-regulating prpC, the enzyme of the *mcs* reaction, can be answered. As becomes clear from the flux control coefficients, the only significant control under biosynthetic conditions is exerted by *mcs* (Figure 7.8, lower left). Hence, only an up-regulation of *mcs*

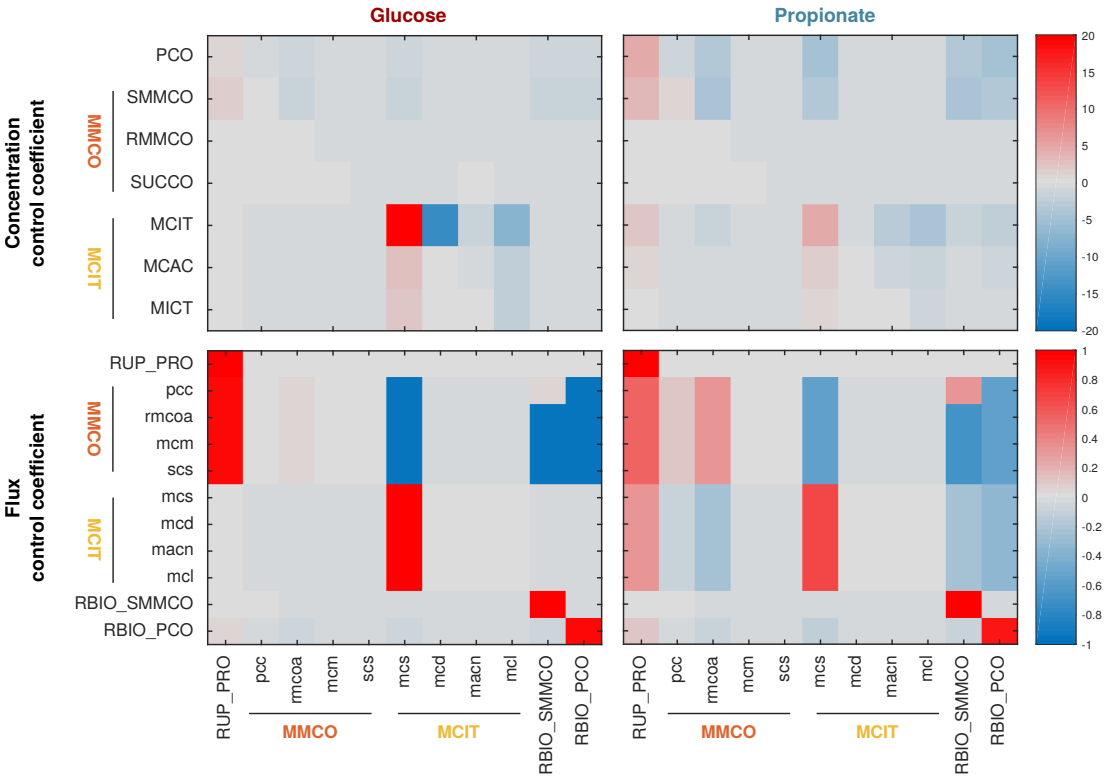


Figure 7.8: Control coefficients from metabolic control analysis. The color coded (red - positive control, blue - negative control, intensity of the color encodes strength) flux control coefficients (lower panel) and concentration control coefficients (upper panel) are shown. The reactions labeled on the x-axis thereby exert the control on the components shown on the y-axis. The respective response coefficients are shown in Appendix A.34.

can significantly increase the steady state catabolic flux as required when propionate is to be used as a carbon source. The flux through MMCO reactions is only passively following the propionate uptake rate (Figure 7.8, first column, *RUP_PRO*). Under catabolic conditions (Figure 7.8, lower right) the control shifts also to the reactions of the MMCO pathway, consistent with the pathway's compensating function as described before [51, 173]. The *mcs* reaction, not surprisingly, executes also the strongest control on the concentration levels of the toxic intermediate methylcitrate (Figure 7.8, MCIT, upper left), highlighting the pathway's role in embanking those concentrations.

7.2.3 Discussion

Professional high-capacity catabolism *vs.* flexibility for rapid detoxification

The presented analyses characterize the MCIT pathway as a high-capacity catabolic pathway that exploits thermodynamically beneficial reaction steps, which operate below saturation in physiological scenarios. The MMCO pathway acts as a flexible pathway capable of rapid PCO detoxification but also of biosynthetic tasks. This flexibility is facilitated by a fine-tuned thermodynamic balance at the cost of an upper catabolic flux capacity limit.

This distinct functionalities render it beneficial for the cell to retain both pathways despite their redundancy in channeling carbon from PCO to the central carbon metabolism. Implementing both in one pathway would reduce the efficiency of either functionality, as a high forward flux capacity with a steep thermodynamic gradient contradicts the ability to quickly invert the flux direction.
Further metabolic tasks of each pathway could be seen as downstream of the described functionalities. The constitutively expressed MMCO pathway also supplies precursors for virulence relevant lipids, which are required under all environmental conditions. The MCIT pathway, as an additional metabolic advantage, directly feeds pyruvate into lower glycolysis, thereby skipping potentially expensive gluconeogenetic reactions.
The plasticity of this crucial lipid biosynthetic and catabolic node has recently been shown to be relevant in eukaryotes, with implications on human metabolic diseases [206], arguing for a general metabolic strategy.

Implications on treatment strategies

Eoh and Rhee [51] as well as Savvi and coworkers [173] show that the MMCO pathway can compensate for the loss of MCIT pathway functionality in mutants lacking *mcl* activity in terms of survival but cannot support growth in an equally efficient manner. However, Van der Ven *et al.* [202] and Upton and coworkers [198] pointed out that the inhibition of the upstream enzyme prpC can reverse this effect, implying that the growth defect is due to the accumulation of toxic MCIT intermediates rather than due to a less efficient metabolic processing of carbon through the MMCO pathway. This argument can be supported by the extremely negative $\Delta_r G^0$ of the corresponding *mcs* reaction, which results in rather high concentrations of the toxic product methylcitrate even if the downstream flux is blocked. Consequently, the optimum bactericidal strategy would be to inhibit both a downstream MCIT reaction such as *mcl* as well as the first enzyme of the MMCO pathway, thereby increasing toxic intermediate levels and abrogating the buffering flux bypass completely.

Dynamics and quantitative data make the difference

The thermodynamic analyses clearly highlight the condition dependent nature of protein cost optimality: The cheapest pathway in one condition might be out-competed by an alternative route under different conditions. Despite the usefulness of optimum protein cost considerations in previous applications, they seem insufficient to fully explain the interplay between the two pathways in question here. One reason could be the low level of interdependency between the reactions on the metabolite level, *i.e.* they employ an almost disjunct set of metabolites, resulting in many degrees of freedom for the optimization. This means that metabolite levels can be tuned independently to achieve a low overall cost due to backward flux (*i.e.* the thermodynamic term, yellow bars in Figure 7.3) and only the unknown kinetic parameters elevate the cost in a way that might not be representative of the saturation state *in vivo*. Accordingly, these results argue for the importance of enzyme kinetics for the understanding of pathway flux.
Furthermore, protein distribution and cost optimality approaches are limited by their focus on time-scales that allow for equilibration of the metabolism for describing the evolutionary benefit of reducing resources and increasing metabolic outcome [48, 57, 92, 214]. However, as *M. tuberculosis* invests only a small fraction of its resources in proteins (see Chapter 4), the evolutionary pressure of reducing protein cost might not be high for the bacteria. In addition, the cells might have other objectives for pathway functionality that are not detectable by optimizing for a constant condition, especially for bacteria subjected to variable environmental

conditions like *M. tuberculosis* has to face in the human host. In this context, multi-objective Pareto optimality has been found to describe the overall behavior of systems well [48, 179]. Here, a distinct, temporal objective is added, describing metabolic non-steady state optimality. The transient behavior in response to sudden metabolic challenges has proven to be essential for the system in question here.
The results also highlight the importance of concentration constraints based on metabolite measurements: Using only free metabolite concentrations within physiologically sensitive bounds, the protein cost optimization would prefer the MCIT pathway, entirely based on its lower weight enzymes. The actual intracellular concentrations are far more strict constraints on the feasible flux space, which underlines the importance of high quality metabolomics data to understand the underlying mechanisms.

Kinetic models in the -omics era

Despite the surely occurring flaws of modeling approaches – here for example the failure to resolve the dynamics of propionate decay with high confidence – mechanistic conclusions can be drawn from carefully calibrated models. However, they can only yield predictions up to the level of accuracy and coverage of the data. Based on that, they can identify gaps in the knowledge about the system in an unbiased manner. With the help of the model, parts of the system are still ill-defined can be pinpointed. This can, for example, be done by identifying parameter non-identifiabilities which were present for the majority of the model parameters, despite the high number of data points included. Targeted follow-up experiments can then resolve those knowledge gaps.
The availability of high quality metabolomics data allows for the integration of thermodynamic principles into kinetic metabolic models, which has been shown to be very insightful here. However, despite the availability of several kinetic-thermodynamic rate laws (see Methods Section 2.2.3), the vast majority of published kinetic metabolic models does not take into account thermodynamics.

IV Conclusions and Outlook

8. Conclusions

The previous Chapters presented the major findings and discussed them individually with regard to their accuracy, biological recovery, possible sources of error, potential improvements and others. This concluding Chapter takes up these threads and orients the presented results in the larger scientific context. The novel findings and their implications are highlighted, their contribution to open scientific problems is explained.

8.1 Strategies for modeling large metabolic networks

Driven by the increasing availability of omics data, systems biological modeling approaches strive for larger coverage of cellular processes while increasing mechanistic accuracy. Especially in metabolic modeling, previous studies either cover a small part of metabolism in great detail [31, 94, 145, 201] or the entire metabolic network on a stoichiometric level [23]. Increasing coverage of the former and mechanistic detail of the latter approaches are frontiers in the systems biology of cellular metabolism, which are also tackled in the work presented here.

The right granularity for modeling

Besides their predictive power, mathematical models can be seen as a formalization of the current state of knowledge on a given system. All available data can be integrated in and form the basis of this formalization. Accordingly, the resolutions of model and data have to match. If for example only a pool of metabolites can be measured together, modeling efforts will not be able to reveal the properties of reactions inter-converting between the pooled metabolites. However, predictions on the overall behavior of this metabolite pool can be drawn with high accuracy. From this point of view, modeling efforts also identify missing knowledge, which can in no case be extracted from the available data. At the same time, they can be used to propose meaningful experiments, capable of specifically filling the knowledge gaps.

In this thesis, the system in question – the central carbon metabolism in mycobacteria – is interrogated at different levels of granularity, exemplifying the interdependency of models and their data. Specifically, in Chapter 5.2 steady state metabolite concentrations were used to identify thermodynamic bottlenecks in glycolysis in the overall metabolic context, without the need for specific kinetic information. However, also the bias of missing metabolite data was highlighted, locating boundaries of the information available on the system.

In Chapter 6.2, protein data and enzyme kinetic formulations were included to detect regulatory points from the steady state data. With the assumption of the cell's striving for minimality, they could be identified without pinpointing the values for the kinetic parameters in a multi-fitting approach. While the information density was not high enough to predict the kinetic behavior of the metabolic network, dependencies among the kinetic parameters could be disclosed, to which all sets of feasible kinetic parameters will have to obey (see for example Figure 6.3).

Similar statements could be made for the full dynamic model of propionate detoxification in Chapter 7: Despite the high coverage of time-resolved data, many of the model's parameters

remained non-identifiable after the model calibration. The conclusions that could be drawn nevertheless, such as information on the catalytic efficiencies or on the kinetic behavior common to all feasible parameterizations, is then soundly backed by the data (see Figures 7.6 - 7.8).
Together these findings show how to exploit available data, while revealing and adhering to their intrinsic limitations. The strength of the mathematical modeling efforts is thereby to mechanistically back the data and reveal causal relationships from them, such as bottleneck reactions (Chapter 5.2), required regulatory points (Chapter 6.2) and reasons for flux capacity bounds (Chapter 7).

The context makes the difference

Metabolism is densely interlinked to all other cellular functionalities. Cell growth, DNA replication, transcription, translation, signaling and cellular maintenance require functional biomass as well as energy, both supplied by metabolism. All these processes also feed back on their fueling fluxes in a condition dependent manner. Accordingly, metabolism is highly context dependent. In addition to those supply-demand constraints, metabolism also has to obey basic principles of thermodynamics and chemical transformations. Formalizing all those aspects into mathematical models yields highly constrained systems. On one hand, this increases biological accuracy and therefore predictive power of models. On the other hand, model calibration becomes a highly complex problem, requiring to search a high-dimensional parameter space that is strongly and nonlinearly constrained.
The results presented here show how to benefit from information about the network context and provide novel ways to tackle the practical problems arising from contextualization. Biomass backtracking provides a method to integrate knowledge from well curated genome-scale models as a context for small kinetic models (Chapter 5.1). Those small models are easier to calibrate and simulate, as is shown for example in Chapter 7 for mycobacterial propionate catabolism, but still include information on the surrounding network. Thereby also sufficient flux for the generation of biomass is guaranteed.
Furthermore, context dependencies of the intracellular thermodynamic landscape (Chapter 5.2), metabolic regulation (Chapter 6.2) and the cost of protein production (Chapter 7) were described, highlighting the importance of considering cellular contexts in metabolic modeling efforts.

On steady states and dynamics

Most metabolite data are measured in steady state, as the very fast transient time-scales of metabolic reactions are difficult to access experimentally. In the dynamic metabolite data presented here, many metabolites reached a new steady state after less than one minute (see Figure 7.5 and Appendix A.7), a very short time span for complex experimental measurements. Proteomes change on much slower time-scales and are hence measured in larger intervals corresponding to distinct metabolic steady states.
For those reasons, steady state methods are a major focus of this work, yielding widely applicable methods for similar data. While important insights can be gained from the system neglecting enzyme kinetics, such as the thermodynamic bottlenecks in Chapter 5.2, kinetic terms can additionally be included in steady state models to improve performance. In existing methods [29, 59] as well as in the results presented in Chapter 6.2, the consideration of enzyme saturation via suitable kinetic laws increases the level of mechanistic detail compared to merely stoichiometric or thermodynamic methods. For unknown parameters, sampling approaches are frequently used, as opposed to the multiple start point optimization presented here.

Steady state problems usually have a set of solutions, rather than an unique one, resulting in parameter non-identifiabilities and dependent parameters (see Figure 6.3). These information can be exploited for the calibration of the time-resolved behavior of the model to dynamic data: All feasible parameterization will have to lie within the set of solutions to the steady state problem.
However, if kinetic data can be obtained, it will likely contain more information than steady state measurements. As shown in Chapter 7, some mechanisms only reveal themselves on the dynamic time-scale. While considering thermodynamics and enzyme costs in steady state were not able to explain pathway preferences, a model fitted to time-resolved metabolite data succeeded. Accordingly, if time resolved data are available, they should be exploited, but a lot can be learned from steady state data and modeling already.

8.2 Metabolic flexibility confirmed

Together all described methods and results add to the understanding of *M. tuberculosis* biology in the pathological context. Especially the open question of how the bacteria reroute their flux in changing environmental conditions was addressed. The results likely represent general strategies to deal with harsh nutritional and environmental conditions
Backup pathways were proven to be an effective way to overcome thermodynamic limitations when efficient catabolism is required, while retaining the potential to react quickly to sudden metabolic challenges, using the example of propionyl-CoA catabolism (Chapter 7). Both functionalities implemented in one pathway would lower either efficiency, as was observed for the glycolytic/gluconeogenetic pathway in Chapter 5.2. Glycolysis was identified there as thermodynamic bottleneck in both directions, based on the measured metabolite concentrations. Why there is no set of parallel pathways for glycolysis and gluconeogenesis can be hypothesized: There might not be a sufficiently good alternative, as glycolysis is generally considered a minimal pathway [135]. Another explanation might be that very little gluconeogenesis is required *in vivo*, potentially due to sugars scavenged from the host, as has been shown to be relevant during persistence [114].
The high level of conservation of protein amounts (Figure 4.2 B) and biomass leaks (Figure 5.4 C and Appendix A.10) between the conditions indicated that only a small, targeted set of changes is executed by the cells to transit from growth on one nutrient to another. While protein changes cannot explain this transition fully, their combination with metabolite concentrations in Chapter 6.2 revealed possible points of regulation, which can now be targeted by specific experiments.

9. Outlook

As a final part of this work, an outlook is given in the following Chapter, covering ongoing work and future implications of the presented findings.

9.1 Towards modeling larger metabolic networks in more detail

Integration of all available data

With the era of big data advancing, modeling efforts will aim to include more and more diverse datasets to confine the metabolic solution space further and further [157]. Especially for large scale or even whole cell modeling, metabolism gives context for other cellular processes and at least for simpler organisms, cellular function is to a large extent governed by the metabolic state. Accordingly, the need for large, high-accuracy metabolic models will increase. The presented methods may find application in the construction and calibration of such models. Especially backtracking is a handy tool for consistently generating sub-models that are easily integrable into larger frameworks and well parameterizable, fulfilling the needs of modular modeling approaches. The method also allows to transmit information on increasingly available patient proteomes and metabolomes via sophisticated FBA methods to detailed dynamic models and can hence have applications in personalized medicine [33].

Towards a thermodynamic-kinetic CCMtb

Based on the results of the thermodynamic-kinetic model in Chapter 7, ongoing efforts are elated about the expansion of the method to the entire mycobacterial central carbon metabolism, thereby fully exploiting the dynamic metabolite data. However, in a dynamic model of that size the pitfalls of full thermodynamic-kinetic equations become evident. They are highly nonlinear, which results in an expansion of the derivative matrices used for numerical computation and extremely long simulation times. Optimization routines accordingly struggle with the uncomfortable objective function as well as with the high dimensional parameter space, covering kinetic parameters, initial conditions, uptake and secretion parameters, Gibbs energies and error model parameters. On the way to a calibrated model, automatized model and data setup scripts have proven useful to ensure mathematical consistency. Also a separate biomass flux for each backtracked compound improved numeric stability.

While the model is far from finished, first simulations in good accordance with the data can be obtained (example in Appendix A.35). An interesting aspect is the multi-objective of the model to comply with dynamic experimental data on one hand, but also arrive at a valid steady state. A steady state pre-fit, similar to the workflow in Chapter 6.2, has proven useful in constraining the feasible parameter space. Another focus of ongoing work is the improvement of algorithms capable of estimating parameters in steady state systems efficiently.

Especially in the context of genome-scale metabolic data becoming available, this ongoing study exemplifies the need for an automated generation of dynamic models, which does not require detailed measurement of each involved enzymes kinetics and tackles the issues arising from the large size of the system.

Frontiers in modeling large metabolic network

Kinetic modeling on the genome-scale is largely limited by the availability of efficient parameter estimation techniques. Computational costs of estimating hundreds of parameters in an highly nonlinear parameter space are humongous and fitting routines do not converge well. Especially, the finding of beneficial initial points for the optimization would ease the task.
In addition also high resolution data is only available for a small subset of the entire metabolic system. The data used in this work have an exceptionally high coverage and (time-)resolution – more remote pathways in other organisms cannot be studied equally well due to the lack of available data. Nevertheless, most of the model parameters stay non-identifiable, arguing again for a close coordination of experimental and modeling efforts. The mathematical models can thereby identify measurement scenarios (time-points or perturbations) that will contain maximum information. Even though, omics data are created by the hour and their quality is steadily improving, there is a lack of data tailored to modeling. This holds also for measured kinetic parameters of enzymes, which are key to kinetic models. For example, for none of the enzymes of the model organism *E. coli* a full set of kinetic parameters and educt concentrations is available from literature [46]. Despite the fact that almost all data can be exploited by modeling efforts, they have a much higher power with the right kind of data: Quantitative measurements of the specifically involved compounds and their kinetic parameters.

9.2 Tuberculosis on the verge of extinction?

The World Health Organization has transited from their "Stop TB" to the "End TB" strategy in 2015, counting major successes in the fight against tuberculosis around the world: 43 million lives were saved in the 15 years before, owing to improved diagnosis and treatment, with an 86 % treatment success rate for newly-diagnosed patients. However, they also clearly state that "[n]ew diagnostics, drugs and vaccines will be needed to achieve the targets set in the End TB Strategy" [209].
Global tuberculosis control faces mostly practical problems, such as to guarantee the duly administration of drugs in weak infrastructure countries, the low efficacy of vaccination and the reduced antibiotic impact in HIV-co-infections. Tuberculosis is in this sense different from other fast progressing pandemics. Fighting tuberculosis requires long and expensive public health campaigns that show success only after years or decades.
Speeding things up with more efficient antibiotics and vaccination strategies is therefore a major concern of basic research on *M. tuberculosis* and the relevant host immune mechanisms. The practical problem is to reveal the details and complex mechanisms of *M. tuberculosis*' pathology to understand where and why currently available treatments are failing – and to find novel ways to overcome these obstacles. In this required systemic research, systems biology and computational modeling will lead the way, equipped with on-target, high quality experimental data.

Bibliography

[1] http://bioinformatics.psb.ugent.be/webtools/Venn/.

[2] L. Aaron, D. Saadoun, I. Calatroni, O. Launay, N. Mémain, V. Vincent, G. Marchal, B. Dupont, O. Bouchaud, D. Valeyre, and O. Lortholary. Tuberculosis in HIV-infected patients: A comprehensive review. *Clinical Microbiology and Infection*, 10(5):388–398, 2004.

[3] R. A. Alberty. *Thermodynamics of biochemical reactions*. John Wiley & Sons, 2005.

[4] B. B. Aldridge, M. Fernandez-Suarez, D. Heller, V. Ambravaneswaran, D. Irimia, M. Toner, and S. M. Fortune. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science (New York, N.Y.)*, 335(6064):100–4, jan 2012.

[5] K. Anand, D. Mathur, A. Anant, and L. C. Garg. Structural studies of phosphoglucose isomerase from Mycobacterium tuberculosis H37Rv. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 66(5):490–497, 2010.

[6] J. Anderson, Y. C. Chang, and A. Papachristodoulou. Model decomposition and reduction tools for large-scale networks in systems biology. *Automatica*, 47(6):1165–1174, 2011.

[7] H. W. Aung, S. a. Henry, and L. P. Walker. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Industrial biotechnology (New Rochelle, N.Y.)*, 9(4):215–228, 2013.

[8] J. J. Baker, B. K. Johnson, and R. B. Abramovitch. Slow growth of Mycobacterium tuberculosis at acidic pH is regulated by phoR and host-associated carbon sources. *Molecular Microbiology*, 94(1):56–69, oct 2014.

[9] S. Banerjee, A. Nandyala, R. Podili, V. M. Katoch, and S. E. Hasnain. Comparison of Mycobacterium tuberculosis isocitrate dehydrogenases (ICD-1 and ICD-2) reveals differences in coenzyme affinity, oligomeric state, pH tolerance and phylogenetic affiliation. *BMC biochemistry*, 6(1):20, 2005.

[10] S. Banerjee, A. K. Nandyala, P. Raviprasad, N. Ahmed, and S. E. Hasnain. Iron-dependent RNA-binding activity of Mycobacterium tuberculosis aconitase. *Journal of bacteriology*, 189(11):4046–52, jun 2007.

[11] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–10, may 2011.

[12] A. Bateman et al. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2015.

[13] A. D. Baughn and K. Y. Rhee. Metabolomics of Central Carbon Metabolism in Mycobacterium tuberculosis. pages 1–16, 2014.

[14] D. Beard, S.-d. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83(1):79–86, jul 2002.

[15] D. A. Beard, E. Babson, E. Curtis, and H. Qian. Thermodynamic constraints for biochemical networks. *Journal of theoretical biology*, 228(3):327–33, jun 2004.

[16] D. A. Beard and H. Qian. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLoS ONE*, 2(1):1–4, 2007.

[17] S. a. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3):727–38, jan 2007.

[18] V. Becker, M. Schilling, J. Bachmann, U. Baumann, A. Raue, T. Maiwald, J. Timmer, and U. Klingmüller. Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science (New York, N.Y.)*, 328(5984):1404–1408, 2010.

[19] M. A. Behr, M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science (New York, N.Y.)*, 284(5419):1520–3, may 1999.

[20] D. Beste and J. Peters. Compiling a molecular inventory for Mycobacterium bovis BCG at two growth rates: evidence for growth rate-mediated regulation of ribosome biosynthesis and lipid metabolism. *Journal of Bacteriology*, 187(5):1677–1684, 2005.

[21] D. J. V. Beste, B. Bonde, N. Hawkins, J. L. Ward, M. H. Beale, S. Noack, K. Nöh, N. J. Kruger, R. G. Ratcliffe, and J. McFadden. (13)C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation. *PLoS pathogens*, 7(7):e1002091, jul 2011.

[22] D. J. V. Beste, K. Nöh, S. Niedenführ, T. a. Mendum, N. D. Hawkins, J. L. Ward, M. H. Beale, W. Wiechert, and J. McFadden. (13)C-Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular Mycobacterium tuberculosis. *Chemistry & biology*, pages 1–10, jul 2013.

[23] A. Bordbar, J. M. Monk, Z. a. King, and B. O. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nature reviews. Genetics*, 15(2):107–20, feb 2014.

[24] H. I. M. Boshoff and C. E. Barry. Tuberculosis - metabolism and respiration in the absence of growth. *Nature reviews. Microbiology*, 3(1):70–80, jan 2005.

[25] G. E. Briggs and J. B. S. Haldane. A note on the kinematics of enzyme action. *Biochem. J.*, 19:338–339, 1925.

[26] J. M. Buescher, S. Moco, U. Sauer, and N. Zamboni. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Analytical Chemistry*, 82(11):4403–4412, 2010.

[27] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2008.

[28] J. Carrera, R. Estrela, J. Luo, N. Rai, A. Tsoukalas, and I. Tagkopoulos. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of Escherichia coli. *Molecular systems biology*, 10(7):735, jan 2014.

[29] A. Chakrabarti, L. Miskovic, K. C. Soh, and V. Hatzimanikatis. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology Journal*, 8(9):1043–1057, 2013.

[30] A. Chang, I. Schomburg, S. Placzek, L. Jeske, M. Ulbrich, M. Xiao, C. W. Sensen, and D. Schomburg. BRENDA in 2015: Exciting developments in its 25th year of existence. *Nucleic Acids Research*, 43(D1):D439–D446, 2015.

[31] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, 79(1):53–73, jul 2002.

[32] V. Chelliah, C. Laibe, and N. Le Novère. BioModels Database: A Repository of Mathematical Models of Biological Processes. *Methods in molecular biology (Clifton, N.J.)*, 1021:189–99, jan 2013.

[33] R. Chen and M. Snyder. Systems biology: Personalized medicine for the future? *Current Opinion in Pharmacology*, 12(5):623–628, 2012.

[34] A. Chowdhury, A. R. Zomorrodi, and C. D. Maranas. k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS computational biology*, 10(2):e1003487, feb 2014.

[35] V. Chubukov, L. Gerosa, K. Kochanowski, and U. Sauer. Coordination of microbial metabolism. *Nature Reviews Microbiology*, 12(5):327–340, mar 2014.

[36] V. Chubukov, M. Uhr, L. Le Chat, R. J. Kleijn, M. Jules, H. Link, S. Aymerich, J. Stelling, and U. Sauer. Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. *Molecular systems biology*, 9(709):709, jan 2013.

[37] S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685):537–44, jun 1998.

[38] C. Colijn, A. Brandes, J. Zucker, D. S. Lun, B. Weiner, M. R. Farhat, T.-Y. Cheng, D. B. Moody, M. Murray, and J. E. Galagan. Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS computational biology*, 5(8):e1000489, aug 2009.

[39] S. E. Connor, G. C. Capodagli, M. K. Deaton, and S. D. Pegan. Structural and functional characterization of Mycobacterium tuberculosis triosephosphate isomerase. *Acta Crystallographica Section D: Biological Crystallography*, 67(12):1017–1022, 2011.

[40] G. M. Cook, M. Berney, S. Gebhard, M. Heinemann, R. A. Cox, O. Danilchanka, and M. Niederweis. Physiology of mycobacteria. *Advances in microbial physiology*, 55(09):81–182, 318–9, 2009.

[41] G. M. Cook, K. Hards, C. Vilchèze, T. Hartman, and M. Berney. Energetics of Respiration and Oxidative Phosphorylation in Mycobacteria. *Microbiology spectrum*, 2(3):1–20, jun 2014.

[42] M. W. Covert and B. Ø. Palsson. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *The Journal of biological chemistry*, 277(31):28058–64, aug 2002.

[43] R. A. Cox. Quantitative relationships for specific growth rates and macromolecular compositions of Mycobacterium tuberculosis, Streptomyces coelicolor A3(2) and Escherichia coli B/r: an integrative theoretical approach. *Microbiology (Reading, England)*, 150(Pt 5):1413–26, may 2004.

[44] A. F. Cunningham and C. L. Spreadbury. Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *Journal of bacteriology*, 180(4):801–8, feb 1998.

[45] G. Dantzig, A. Orden, and P. Wolfe. The Generalized Simplex Method for Minimizing a Linear form Under Linear Inequality Restraints. *Pacific Journal of Mathematics*, 5(2):183–195, 1955.

[46] D. Davidi, E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, and R. Milo. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k cat measurements. *Proceedings of the National Academy of Sciences*, 113(12):3401–3406, mar 2016.

[47] L. P. S. De Carvalho, Y. Ling, C. Shen, J. D. Warren, and K. Y. Rhee. On the chemical mechanism of succinic semialdehyde dehydrogenase (GabD1) from Mycobacterium tuberculosis. *Archives of Biochemistry and Biophysics*, 509(1):90–99, 2011.

[48] G. M. de Hijas-Liste, E. Klipp, E. Balsa-Canto, and J. R. Banga. Global dynamic optimization approach to predict activation in metabolic pathways. *BMC systems biology*, 8:1, 2014.

[49] N. K. Dutta and P. C. Karakousis. Latent tuberculosis infection: myths, models, and molecular mechanisms. *Microbiology and molecular biology reviews : MMBR*, 78(3):343–71, sep 2014.

[50] M. Ederer and E. D. Gilles. Thermodynamically feasible kinetic models of reaction networks. *Biophysical journal*, 92(6):1846–57, mar 2007.

[51] H. Eoh and K. Y. Rhee. Methylcitrate cycle defines the bactericidal essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty acids. *Proceedings of the National Academy of Sciences of the United States of America*, 111(13):4976–81, apr 2014.

[52] A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. Ø. Palsson. Reconstruction of biochemical networks in microorganisms. *Nature reviews. Microbiology*, 7(2):129–43, feb 2009.

[53] D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal*, 238(3):781–786, 1986.

[54] D. M. Ferraris, R. Spallek, W. Oehlmann, M. Singh, and M. Rizzi. Structures of citrate synthase and malate dehydrogenase of Mycobacterium tuberculosis. *Proteins*, 83(2):389–94, feb 2015.

[55] R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, 44(D1):D279–D285, 2015.

[56] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London (Series A)*, 222(594-604):309–368, jan 1922.

[57] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences of the United States of America*, 110(24):10039–44, jun 2013.

[58] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo. eQuilibrator–the biochemical thermodynamics calculator. *Nucleic acids research*, 40(Database issue):D770–5, jan 2012.

[59] R. M. T. Fleming, I. Thiele, G. Provan, and H. P. Nasheuer. Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. *Journal of theoretical biology*, 264(3):683–92, jul 2010.

[60] E. Fullam, F. Pojer, T. Bergfors, T. A. Jones, and S. T. Cole. Structure and function of the transketolase from Mycobacterium tuberculosis and comparison with the human enzyme. *Open biology*, 2(1):110026, jan 2012.

[61] S. Gagneux. Host-pathogen coevolution in human tuberculosis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1590):850–9, 2012.

[62] G. Gago, D. Kurth, L. Diacovich, S.-c. Tsai, and H. Gramajo. Biochemical and structural characterization of an essential acyl coenzyme A carboxylase from Mycobacterium tuberculosis. *Journal of bacteriology*, 188(2):477–86, jan 2006.

[63] M. Gengenbacher and S. H. E. Kaufmann. Mycobacterium tuberculosis: Success through dormancy. *FEMS Microbiology Reviews*, 36(3):514–532, 2012.

[64] M. Gengenbacher, S. P. S. Rao, K. Pethe, and T. Dick. Nutrient-starved, non-replicating Mycobacterium tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance of ATP homeostasis and viability. *Microbiology (Reading, England)*, 156(Pt 1):81–7, jan 2010.

[65] M. P. Gerstl, D. E. Ruckerbauer, D. Mattanovich, C. Jungreuthmayer, and J. Zanghellini. Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Scientific reports*, 5:8930, 2015.

[66] R. N. Goldberg, Y. B. Tewari, and T. N. Bhat. Thermodynamics of enzyme-catalyzed reactions–a database for quantitative biochemistry. *Bioinformatics (Oxford, England)*, 20(16):2874–7, nov 2004.

[67] T. A. Gould, H. van de Langemheen, E. J. Muñoz-Elías, J. D. McKinney, and J. C. Sacchettini. Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis. *Molecular microbiology*, 61(4):940–7, aug 2006.

[68] J. E. Griffin, A. K. Pandey, S. A. Gilmore, V. Mizrahi, J. D. McKinney, C. R. Bertozzi, and C. M. Sassetti. Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. *Chemistry & biology*, 19(2):218–27, feb 2012.

[69] J. Hastings, P. de Matos, A. Dekker, M. Ennis, B. Harsha, N. Kale, V. Muthukrishnan, G. Owen, S. Turner, M. Williams, and C. Steinbeck. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic acids research*, 41(Database issue):D456–63, jan 2013.

[70] R. Heinrich and T. a. Rapoport. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European journal of biochemistry / FEBS*, 42(1):89–95, feb 1974.

[71] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical journal*, 92(5):1792–805, mar 2007.

[72] J.-h. S. Hofmeyr. Metabolic control analysis in a nutshell. *International conference on systems biology*, pages 291–300, 2001.

[73] H. G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, 271(14):2905–2922, 2004.

[74] K. Höner Zu Bentrup, A. Miczak, D. L. Swenson, and D. G. Russell. Characterization of activity and expression of isocitrate lyase in Mycobacterium avium and Mycobacterium tuberculosis. *Journal of bacteriology*, 181(23):7161–7, dec 1999.

[75] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, 22(24):3067–74, dec 2006.

[76] A. Hoppe, S. Hoffmann, and H.-G. Holzhütter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC systems biology*, 1:23, jan 2007.

[77] P. C. Hsieh, T. H. Kowalczyk, and N. F. Phillips. Kinetic mechanisms of polyphosphate glucokinase from Mycobacterium tuberculosis. *Biochemistry*, 35(30):9772–81, jul 1996.

[78] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, mar 2003.

[79] F. Hynne, S. Danø, and P. Sørensen. Full-scale model of glycolysis in Saccharomyces cerevisiae. *Biophysical Chemistry*, 94(1-2):121–163, dec 2001.

[80] M. Jain, C. J. Petzold, M. W. Schelle, M. D. Leavell, J. D. Mougous, C. R. Bertozzi, J. a. Leary, and J. S. Cox. Lipidomics reveals control of Mycobacterium tuberculosis virulence lipids via metabolic coupling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(12):5133–5138, 2007.

[81] C. J. Jeffery. Moonlighting proteins: Old proteins learning new tricks, 2003.

[82] P. Jouhten, E. Rintala, A. Huuskonen, A. Tamminen, M. Toivari, M. Wiebe, L. Ruohonen, M. Penttilä, and H. Maaheimo. Oxygen dependence of metabolic fluxes and energy generation of Saccharomyces cerevisiae CEN.PK113-1A. *BMC systems biology*, 2:60, 2008.

[83] H. Kacser and J. A. Burns. The control of flux. *Symposia of the Society for Experimental Biology*, 27:65–104, 1973.

[84] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1):D457–D462, 2016.

[85] R. Kapoor and T. A. Venkitasubramanian. Purification and properties of pyruvate kinase from Mycobacterium smegmatis. *Archives of biochemistry and biophysics*, 225(1):320–30, aug 1983.

[86] J. Karr, J. Sanghvi, D. Macklin, M. Gutschow, J. Jacobs, B. Bolival, N. Assad-Garcia, J. Glass, and M. Covert. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*, 150(2):389–401, jul 2012.

[87] S. H. Kaufmann. How can immunology contribute to the control of tuberculosis? *Nature reviews. Immunology*, 1(1):20–30, oct 2001.

[88] S. H. Kaufmann and W. J. Britton. *Handbook of Tuberculosis: Immunology and Cell Biology*. Wiley-Blackwell, 2008.

[89] S. H. Kaufmann and E. J. Rubin. *Handbook of Tuberculosis: Molecular Biology and Biochemistry.* Wiley-VCH, 2008.

[90] S. M. Keating, B. J. Bornstein, A. Finney, and M. Hucka. SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics (Oxford, England)*, 22(10):1275–7, may 2006.

[91] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science (New York, N.Y.)*, 220(4598):671–80, may 1983.

[92] E. Klipp, R. Heinrich, and H. G. Holzhütter. Prediction of temporal gene expression: Metabolic optimization by re-distribution of enzyme activities. *European Journal of Biochemistry*, 269(22):5406–5413, 2002.

[93] E. Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems Biology. A Textbook.* Wiley-Blackwell, Weinheim, 2nd edition, 2016.

[94] E. Klipp, B. Nordlander, R. Krüger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nature biotechnology*, 23(8):975–82, aug 2005.

[95] R. Koch. Die Ätiologie der Tuberkulose (Nach einem in der physiologischen Gesellschaft zu Berlin am 24. März gehaltenem Vortrage). *Berliner klinische Wochenschrift*, 19(15):221–30, 1882.

[96] O. Kotte, J. B. Zaugg, and M. Heinemann. Bacterial adaptation through distributed sensing of metabolic fluxes. *Molecular systems biology*, 6(355):355, jan 2010.

[97] C. Kreutz, A. Raue, D. Kaschek, and J. Timmer. Profile likelihood in systems biology. *The FEBS journal*, 280(11):2564–71, jun 2013.

[98] A. Kümmel, S. Panke, and M. Heinemann. Systematic assignment of thermodynamic constraints in metabolic network models. *BMC bioinformatics*, 7:512, jan 2006.

[99] W. Lee, B. C. VanderVen, R. J. Fahey, and D. G. Russell. Intracellular Mycobacterium tuberculosis exploits host-derived fatty acids to limit metabolic stress. *The Journal of biological chemistry*, 288(10):6788–800, mar 2013.

[100] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[101] J. M. Lew, A. Kapopoulou, L. M. Jones, and S. T. Cole. TubercuList - 10 years after. *Tuberculosis*, 91(1):1–7, 2011.

[102] N. E. Lewis, K. K. Hixson, T. M. Conrad, J. a. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, K. K. Weitz, R. Eils, R. König, R. D. Smith, and B. Ø. Palsson. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6(390):390, jul 2010.

[103] N. E. Lewis, H. Nagarajan, and B. O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature reviews. Microbiology*, 10(4):291–305, apr 2012.

[104] W. Liebermeister and E. Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical biology & medical modelling*, 3(1):41, jan 2006.

[105] W. Liebermeister, E. Noor, A. Flamholz, D. Davidi, J. Bernhardt, and R. Milo. Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8488–93, jun 2014.

[106] W. Liebermeister, J. Uhlendorf, and E. Klipp. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics (Oxford, England)*, 26(12):1528–34, jun 2010.

[107] J. Limenitakis, R. D. Oppenheim, D. J. Creek, B. J. Foth, M. P. Barrett, and D. Soldati-Favre. The 2-methylcitrate cycle is implicated in the detoxification of propionate in Toxoplasma gondii. *Molecular Microbiology*, 87(4):894–908, 2013.

[108] R. R. Lovewell, C. M. Sassetti, and B. C. VanderVen. Chewing the fat: lipid metabolism and homeostasis during M. tuberculosis infection. *Current Opinion in Microbiology*, 29:30–36, 2016.

[109] C. Ludwig, M. Claassen, a. Schmidt, and R. Aebersold. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Molecular & cellular proteomics : MCP*, 11(3):M111.013987, 2012.

[110] I. Machová, J. Snášel, J. Dostál, J. Brynda, J. Fanfrlík, M. Singh, J. Tarábek, O. Vaněk, L. Bednárová, and I. Pichová. Structural and Functional Studies of Phosphoenolpyruvate Carboxykinase from Mycobacterium tuberculosis. *Plos One*, 10(3):e0120682, 2015.

[111] R. Mahadevan, J. S. Edwards, and F. J. Doyle. Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical journal*, 83(3):1331–40, sep 2002.

[112] A. Makhorin. GNU Linear Programming Kit (GLPK), 2016.

[113] D. W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters, 1963.

[114] J. Marrero, C. Trujillo, K. Y. Rhee, and S. Ehrt. Glucose phosphorylation is required for Mycobacterium tuberculosis persistence in mice. *PLoS pathogens*, 9(1):e1003116, jan 2013.

[115] D. Mathur and L. C. Garg. Functional phosphoglucose isomerase from Mycobacterium tuberculosis H37Rv: Rapid purification with high yield and purity. *Protein Expression and Purification*, 52(2):373–378, 2007.

[116] D. Mathur, G. Malik, and L. C. Garg. Biochemical and functional characterization of triosephosphate isomerase from Mycobacterium tuberculosis H37Rv. *FEMS microbiology letters*, 263(2):229–35, oct 2006.

[117] M. L. Mavrovouniotis. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnology and Bioengineering*, 36(10):1070–1082, 1990.

[118] M. L. Mavrovouniotis. Identification of localized and distributed bottlenecks in metabolic pathways. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 1:275–83, jan 1993.

[119] M. L. Mavrovouniotis. Duality theory for thermodynamic bottlenecks in bioreaction pathways. *Chemical Engineering Science*, 51(9):1495–1507, 1996.

[120] M. D. McKay, R. J. Beckman, and W. J. Conover. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, 1979.

[121] H. Mi, S. Poudel, A. Muruganujan, J. T. Casagrande, and P. D. Thomas. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic acids research*, 44(D1):D336–42, jan 2016.

[122] L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochemische Zeitung*, (49):333–369, 1913.

[123] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. BioNumbers The database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(SUPPL.1):750–753, 2009.

[124] M. M. Moynihan and A. S. Murkin. Cysteine is the general base that serves in catalysis by isocitrate lyase and in mechanism-based inhibition by 3-nitropropionate. *Biochemistry*, 53(1):178–187, 2014.

[125] B. Mukhopadhyay and E. Purwantini. Pyruvate carboxylase from Mycobacterium smegmatis: stabilization, rapid purification, molecular and biochemical characterization and regulation of the cellular level. *Biochimica et biophysica acta*, 1475(3):191–206, jul 2000.

[126] E. J. Muñoz-Elías and J. D. McKinney. Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nature medicine*, 11(6):638–44, jun 2005.

[127] E. J. Muñoz-Elías and J. D. McKinney. Carbon metabolism of intracellular bacteria. *Cellular microbiology*, 8(1):10–22, jan 2006.

[128] E. J. Muñoz-Elías, A. M. Upton, J. Cherian, and J. D. McKinney. Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence. *Molecular microbiology*, 60(5):1109–22, jun 2006.

[129] P. Murima, M. Zimmermann, T. Chopra, F. Pojer, G. Fonti, M. Dal Peraro, S. Alonso, U. Sauer, K. Pethe, and J. D. McKinney. A rheostat mechanism governs the bifurcation of carbon flux in mycobacteria. *Nature communications*, 7:12527, 2016.

[130] M. Nandakumar, C. Nathan, and K. Y. Rhee. Isocitrate lyase mediates broad antibiotic tolerance in Mycobacterium tuberculosis. *Nature communications*, 5:4306, jan 2014.

[131] J. Nelder, R. Mead, B. J. a. Nelder, and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1964.

[132] M. Niederweis. Nutrient acquisition by mycobacteria. *Microbiology (Reading, England)*, 154(Pt 3):679–92, mar 2008.

[133] E. Noor, A. Bar-Even, A. Flamholz, Y. Lubling, D. Davidi, and R. Milo. An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics (Oxford, England)*, 28(15):2037–44, aug 2012.

[134] E. Noor, A. Bar-Even, A. Flamholz, E. Reznik, W. Liebermeister, and R. Milo. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS computational biology*, 10(2):e1003483, feb 2014.

[135] E. Noor, E. Eden, R. Milo, and U. Alon. Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy. *Molecular Cell*, 39(5):809–820, 2010.

[136] E. Noor, A. Flamholz, W. Liebermeister, A. Bar-Even, and R. Milo. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS letters*, 587(17):2772–7, sep 2013.

[137] J. D. Orth, T. M. Conrad, J. Na, J. a. Lerman, H. Nam, A. M. Feist, and B. Ø. Palsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Molecular systems biology*, 7(535):535, jan 2011.

[138] H. Ouellet, J. B. Johnston, and P. R. O. de Montellano. Cholesterol catabolism as a therapeutic target in Mycobacterium tuberculosis. *Trends in microbiology*, 19(11):530–9, nov 2011.

[139] N. R. Pace. The universal nature of biochemistry. *Proceedings of the National Academy of Sciences*, 98(3):805–808, jan 2001.

[140] S. Palacios and J. C. Escalante-Semerena. 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of Salmonella enterica by the PrpR protein. *Microbiology*, 150(11):3877–3887, 2004.

[141] B. Palsson. In silico biology through "omics". *Nature biotechnology*, 20(7):649–50, jul 2002.

[142] A. K. Pandey and C. M. Sassetti. Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11):4376–80, mar 2008.

[143] K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, nov 1901.

[144] S. D. Pegan, K. Rukseree, S. G. Franzblau, and A. D. Mesecar. Structural Basis for Catalysis of a Tetrameric Class IIa Fructose 1,6-Bisphosphate Aldolase from Mycobacterium tuberculosis. *Journal of Molecular Biology*, 386(4):1038–1053, 2009.

[145] E. Petelenz-Kurdziel, C. Kuehn, B. Nordlander, D. Klein, K.-K. Hong, T. Jacobson, P. Dahl, J. Schaber, J. Nielsen, S. Hohmann, and E. Klipp. Quantitative Analysis of Glycerol Accumulation, Glycolysis and Growth under Hyper Osmotic Stress. *PLoS Computational Biology*, 9(6):e1003084, jun 2013.

[146] W. H. Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[147] N. D. Price, I. Famili, D. A. Beard, and B. Ø. Palsson. Extreme pathways and Kirchhoff's second law. *Biophysical journal*, 83(5):2879–82, nov 2002.

[148] S. Prisic, S. Dankwa, D. Schwartz, M. F. Chou, J. W. Locasale, C.-M. Kang, G. Bemis, G. M. Church, H. Steen, and R. N. Husson. Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7521–7526, 2010.

[149] C. E. Quartararo and J. S. Blanchard. Kinetic and chemical mechanism of malate synthase from Mycobacterium tuberculosis. *Biochemistry*, 50(32):6879–87, aug 2011.

[150] C. E. Quartararo, S. Hazra, T. Hadi, and J. S. Blanchard. Structural, kinetic and chemical mechanism of isocitrate dehydrogenase-1 from mycobacterium tuberculosis. *Biochemistry*, 52(10):1765–1775, 2013.

[151] O. Radulescu, A. N. Gorban, A. Zinovyev, and V. Noel. Reduction of dynamical biochemical reactions networks in computational biology. *Frontiers in genetics*, 3(July):131, jan 2012.

[152] P. C. Ramsaywak, G. Labbé, S. Siemann, G. I. Dmitrienko, and J. G. Guillemette. Molecular cloning, expression, purification, and characterization of fructose 1,6-bisphosphate aldolase from Mycobacterium tuberculosis–a novel Class II A tetramer. *Protein expression and purification*, 37(1):220–8, sep 2004.

[153] S. Rao, A. van der Schaft, K. van Eunen, B. M. Bakker, and B. Jayawardhana. A model reduction method for biochemical reaction networks. *BMC systems biology*, 8:52, jan 2014.

[154] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics (Oxford, England)*, 25(15):1923–9, aug 2009.

[155] A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelker, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLoS ONE*, 8(9):e74335, 2013.

[156] A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser, W. Mader, T. Heinemann, J. Hasenauer, M. Schilling, T. Höfer, E. Klipp, F. Theis, U. Klingmüller, B. Schöberl, and J. Timmer. Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21):3558–3560, 2014.

[157] J. L. Reed. Shrinking the metabolic solution space using experimental datasets. *PLoS computational biology*, 8(8):e1002662, jan 2012.

[158] K. Y. Rhee, L. P. S. de Carvalho, R. Bryk, S. Ehrt, J. Marrero, S. W. Park, D. Schnappinger, A. Venugopal, and C. Nathan. Central carbon metabolism in Mycobacterium tuberculosis: an unexpected frontier. *Trends in microbiology*, 19(7):307–14, jul 2011.

[159] R. A. Rienksma, M. Suarez-Diez, L. Spina, P. J. Schaap, and V. A. P. Martins Dos Santos. Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets. *Seminars in immunology*, 26(6):610–622, oct 2014.

[160] K. H. Rohde, R. B. Abramovitch, and D. G. Russell. Mycobacterium tuberculosis invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell host & microbe*, 2(5):352–64, nov 2007.

[161] J. M. Rohwer and J.-H. S. Hofmeyr. Identifying and characterising regulatory metabolites with generalised supply-demand analysis. *Journal of theoretical biology*, 252(3):546–54, jun 2008.

[162] A. K. Roos, C. E. Andersson, T. Bergfors, M. Jacobsson, A. Karlen, T. Unge, T. A. Jones, and S. L. Mowbray. Mycobacterium tuberculosis ribose-5-phosphate isomerase has a known fold, but a novel active site. *Journal of Molecular Biology*, 335(3):799–809, 2004.

[163] A. K. Roos, E. Burgos, D. J. Ericsson, L. Salmon, and S. L. Mowbray. Competitive inhibitors of Mycobacterium tuberculosis ribose-5-phosphate isomerase B reveal new information about the reaction mechanism. *Journal of Biological Chemistry*, 280(8):6416–6422, 2005.

[164] A. K. Roos, S. Mariano, E. Kowalinski, L. Salmon, and S. L. Mowbray. d-Ribose-5-Phosphate Isomerase B from Escherichia coli is Also a Functional d-Allose-6-Phosphate Isomerase, While the Mycobacterium tuberculosis Enzyme is Not. *Journal of Molecular Biology*, 382(3):667–679, 2008.

[165] J. Ross and A. P. Arkin. Complex systems: from chemistry to systems biology. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6433–6434, 2009.

[166] S. Rossell, C. C. Van Der Weijden, A. L. Kruckeberg, B. M. Bakker, and H. V. Westerhoff. Hierarchical and metabolic regulation of glucose influx in starved Saccharomyces cerevisiae. *FEMS Yeast Research*, 5(6-7):611–619, 2005.

[167] D. G. Russell, H. C. Mwandumba, and E. E. Rhoades. Mycobacterium and the coat of many lipids. *Journal of Cell Biology*, 158(3):421–426, 2002.

[168] I. Santi, N. Dhar, D. Bousbaine, Y. Wakamoto, and J. D. McKinney. Single-cell dynamics of the chromosome replication and cell division cycles in mycobacteria. *Nature Communications*, 4(May):2470, 2013.

[169] M. J. Sartain, D. L. Dick, C. D. Rithner, D. C. Crick, and J. T. Belisle. Lipidomic analyses of Mycobacterium tuberculosis based on accurate mass measurements and the novel "Mtb LipidDB". *Journal of lipid research*, 52(5):861–872, 2011.

[170] C. M. Sassetti, D. H. Boyd, and E. J. Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1):77–84, mar 2003.

[171] C. M. Sassetti and E. J. Rubin. Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12989–94, oct 2003.

[172] U. Sauer and B. J. Eikmanns. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS microbiology reviews*, 29(4):765–94, sep 2005.

[173] S. Savvi, D. F. Warner, B. D. Kana, J. D. McKinney, V. Mizrahi, and S. S. Dawes. Functional characterization of a vitamin B12-dependent methylmalonyl pathway in Mycobacterium tuberculosis: implications for propionate metabolism during growth on fatty acids. *Journal of bacteriology*, 190(11):3886–95, jun 2008.

[174] D. Schnappinger, S. Ehrt, M. I. Voskuil, Y. Liu, J. A. Mangan, I. M. Monahan, G. Dolganov, B. Efron, P. D. Butcher, C. Nathan, and G. K. Schoolnik. Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages. *The Journal of Experimental Medicine*, 198(5):693–704, sep 2003.

[175] I. Schomburg, A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer, and D. Schomburg. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic acids research*, 41(Database issue):D764–72, jan 2013.

[176] O. Schubert, J. Mouritsen, C. Ludwig, H. Röst, G. Rosenberger, P. Arthur, M. Claassen, D. Campbell, Z. Sun, T. Farrah, M. Gengenbacher, A. Maiolica, S. Kaufmann, R. Moritz, and R. Aebersold. The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. *Cell Host & Microbe*, 13(5):602–612, 2013.

[177] O. T. Schubert, C. Ludwig, M. Kogadeeva, M. Zimmermann, G. Rosenberger, M. Gengenbacher, L. C. Gillet, B. C. Collins, H. L. Röst, S. H. E. Kaufmann, U. Sauer, and R. Aebersold. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of Mycobacterium tuberculosis. *Cell host & microbe*, 18(1):96–108, jul 2015.

[178] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular systems biology*, 3(119):119, 2007.

[179] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional Optimality of Microbial Metabolism. *Science*, 336(2012):601–604, 2012.

[180] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10(8):747–747, aug 2014.

[181] R. Serban and a. C. Hindmarsh. CVODES: the Sensitivity-Enabled ODE Solver in SUNDIALS. *ACM Transactions on Mathematical Software*, 5(September):1–18, 2003.

[182] L. F. Shampine and M. W. Reichelt. The MATLAB ODE Suite. *SIAM Journal on Scientific Computing*, 18(1):1–22, 1997.

[183] L. Shi, C. D. Sohaskey, C. Pfeiffer, P. Datta, M. Parks, J. McFadden, R. J. North, and M. L. Gennaro. Carbon flux rerouting during Mycobacterium tuberculosis growth arrest. *Molecular microbiology*, 78(5):1199–215, dec 2010.

[184] V. K. Singh and I. Ghosh. Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets. *Theoretical biology & medical modelling*, 3:27, jan 2006.

[185] C. V. Smith, C.-c. Huang, A. Miczak, D. G. Russell, J. C. Sacchettini, and K. Höner zu Bentrup. Biochemical and structural studies of malate synthase from Mycobacterium tuberculosis. *The Journal of biological chemistry*, 278(3):1735–43, jan 2003.

[186] N. J. Stanford, T. Lubitz, K. Smallbone, E. Klipp, P. Mendes, and W. Liebermeister. Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS ONE*, 8(11), 2013.

[187] G. R. Stewart, B. D. Robertson, and D. B. Young. Tuberculosis: a problem with persistence. *Nature reviews. Microbiology*, 1(2):97–105, nov 2003.

[188] L. Tailleux, S. J. Waddell, M. Pelizzola, A. Mortellaro, M. Withers, A. Tanne, P. R. Castagnoli, B. Gicquel, N. G. Stoker, P. D. Butcher, M. Foti, and O. Neyrolles. Probing host pathogen cross-talk by transcriptional profiling of both Mycobacterium tuberculosis and infected human dendritic cells and macrophages. *PloS one*, 3(1):e1403, jan 2008.

[189] B. H. ter Kuile and H. V. Westerhoff. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS letters*, 500(3):169–71, jul 2001.

[190] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European journal of biochemistry / FEBS*, 267(17):5313–29, sep 2000.

[191] N. R. Thanky, D. B. Young, and B. D. Robertson. Unusual features of the cell cycle in mycobacteria: Polar-restricted growth and the snapping-model of cell division. *Tuberculosis*, 87(3):231–236, 2007.

[192] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1):93–121, 2010.

[193] J. Tian, R. Bryk, M. Itoh, M. Suematsu, and C. Nathan. Variant tricarboxylic acid cycle in Mycobacterium tuberculosis: identification of alpha-ketoglutarate decarboxylase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10670–5, jul 2005.

[194] J. Tian, R. Bryk, S. Shi, H. Erdjument-Bromage, P. Tempst, and C. Nathan. Mycobacterium tuberculosis appears to lack alpha-ketoglutarate dehydrogenase and encodes pyruvate dehydrogenase in widely separated genes. *Molecular microbiology*, 57(3):859–68, aug 2005.

[195] K. Tummler, C. Kühn, and E. Klipp. Dynamic metabolic models in context: biomass backtracking. *Integr. Biol.*, 7(8):940–951, 2015.

[196] K. Tummler, T. Lubitz, M. Schelker, and E. Klipp. New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. *The FEBS journal*, pages 1–23, sep 2013.

[197] A. K. Tyagi, F. A. Siddiqui, and T. A. Venkitasubramanian. Studies on the purification and characterization of malate dehydrogenase from Mycobacterium phlei. *Biochimica et biophysica acta*, 485(2):255–67, dec 1977.

[198] A. M. Upton and J. D. McKinney. Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. *Microbiology (Reading, England)*, 153(Pt 12):3973–82, dec 2007.

[199] K. Van Eunen, J. Bouwman, A. Lindenbergh, H. V. Westerhoff, and B. M. Bakker. Time-dependent regulation analysis dissects shifts between metabolic and gene-expression regulation during nitrogen starvation in baker's yeast. *FEBS Journal*, 276(19):5521–5536, 2009.

[200] K. van Eunen, J. a. L. Kiewiet, H. V. Westerhoff, and B. M. Bakker. Testing biochemistry revisited: How in vivo metabolism can be understood from in vitro enzyme kinetics. *PLoS Computational Biology*, 8(4), 2012.

[201] J. H. van Heerden, M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. M. Bollen, R. Planqué, J. Hulshof, T. G. O'Toole, S. A. Wahl, and B. Teusink. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science (New York, N.Y.)*, 343(6174):1245114, feb 2014.

[202] B. C. VanderVen, R. J. Fahey, W. Lee, Y. Liu, R. B. Abramovitch, C. Memmott, A. M. Crowe, L. D. Eltis, E. Perola, D. D. Deininger, T. Wang, C. P. Locher, and D. G. Russell. Novel Inhibitors of Cholesterol Degradation in Mycobacterium tuberculosis Reveal How the Bacterium's Metabolism Is Constrained by the Intracellular Environment. *PLOS Pathogens*, 11(2):e1004679, 2015.

[203] T. Wagner, M. Bellinzoni, A. Wehenkel, H. M. O'Hare, and P. M. Alzari. Functional plasticity and allosteric regulation of α-ketoglutarate decarboxylase in central mycobacterial metabolism. *Chemistry & biology*, 18(8):1011–20, aug 2011.

[204] S. Watanabe, M. Zimmermann, M. B. Goodwin, U. Sauer, C. E. Barry, and H. I. Boshoff. Fumarate reductase activity maintains an energized membrane in anaerobic Mycobacterium tuberculosis. *PLoS pathogens*, 7(10):e1002287, oct 2011.

[205] H. A. Watkins, M. Yu, and E. N. Baker. Cloning, expression, purification and preliminary crystallographic data for Rv3214 (EntD), a predicted cofactor-dependent phosphoglycerate mutase from Mycobacterium tuberculosis. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 61(8):753–755, 2005.

[206] E. Watson, V. Olin-Sandoval, M. J. Hoy, C.-H. Li, T. Louisse, V. Yao, A. Mori, A. D. Holdorf, O. G. Troyanskaya, M. Ralser, and A. J. Walhout. Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in C.elegans. *eLife*, 5:1–21, 2016.

[207] R. K. Wierenga, E. G. Kapetaniou, and R. Venkatesan. Triosephosphate isomerase: A highly evolved biocatalyst. *Cellular and Molecular Life Sciences*, 67(23):3961–3982, 2010.

[208] J. a. H. Wodke, J. Puchałka, M. Lluch-Senar, J. Marcos, E. Yus, M. Godinho, R. Gutiérrez-Gallego, V. a. P. M. dos Santos, L. Serrano, E. Klipp, and T. Maier. Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Molecular systems biology*, 9(653):653, jan 2013.

[209] World Health Organization. WHO Global tuberculosis report 2015. Technical report, 2015.

[210] K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, and T. Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics (Oxford, England)*, 26(12):i255–60, jun 2010.

[211] D. B. Young, I. Comas, and L. P. S. de Carvalho. Phylogenetic analysis of vitamin B12-related metabolism in Mycobacterium tuberculosis. *Frontiers in Molecular Biosciences*, 2(March):6, 2015.

[212] N. Zamboni, A. Kümmel, and M. Heinemann. anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data. *BMC bioinformatics*, 9:199, jan 2008.

[213] J. Zanghellini, D. E. Ruckerbauer, M. Hanscho, and C. Jungreuthmayer. Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology Journal*, 8(9):1009–1016, 2013.

[214] A. Zaslaver, A. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. Surette, and U. Alon. Just-in-time transcription program in metabolic pathways. *Nature genetics*, 36(5):486–491, 2004.

[215] J. Zheng, L. Liu, B. Liu, and Q. Jin. Phosphoproteomic analysis of bacillus Calmette-Guérin using gel-based and gel-free approaches. *Journal of Proteomics*, 126:189–199, 2015.

Acknowledgements

This part of the thesis is the one that I was most looking forward to write: Wholehearted thanks to the many contributors, that supported me during the work on this thesis.
First and foremost, I want to thank ***Edda Klipp***, who was a wonderful, always available and caring supervisor. What I learned during the last (more than) five years, is only due to her support and challenges.
Many thanks belong to many collaborators and co-workers which have been nothing less than essential. First of all, again, to ***Clemens Kühn***, who was a work companion with a sharp scientific eye, splendid pragmatism and humor. On the experimental side, I had the pleasure of being part of a – frequently propagated, but rarely observed – close and cheerful collaboration. I want to thank ***Uwe Sauer***, ***Michi Zimmermann*** and ***Olga Schubert***, for the glimpse on the forefront of experimental biology and all the insightful discussions. Many thanks also belong to other experimental collaborators within the SystemTb project: ***Iva Pichova***'s team from Prague and the group of ***Menico Rizzi*** from Vercelli. For the time I worked at the Weizmann Institute I want to thank the hospitable group of ***Ron Milo***, especially ***Elad Noor***, ***Dan Davidi***, ***Uri Barenholz*** and occasionally ***Wolfram Liebermeister*** for the introduction of thermodynamics into my life.
I also have the pleasure to thank the whole group of theoretical biophysics, for being a bunch of wonderful, passionate and immensely clever people. I enjoyed every day of working with them, but also all the off-topics that they introduced me to: Music, extreme sports, brewing and so many others. Special thanks are addressed to ***Max Schelker*** and the spirit of Andreas Raue for numerous hours of support with the D2D tool. For everything and nothing in particular, I want to thank ***Friedemann Uschner*** and ***Jens Hahn***, who became very close friends. I want to thank ***Ivo Maintz*** and ***Sabine Wagnitz*** for always caring about the practical things and making life easy for me. I also want to express my gratitude to all the sedulous proofreaders, especially to the Jungsbüro, to ***Marcus Krantz*** and ***Wolfgang Giese***.
I further want to thank all the funding sources that supported me over the last years: The European tuberculosis project SytemTb, the Computational Systems Biology graduate school, the DAAD, the Minerva Foundation as well as the Humboldt-Universität zu Berlin. I also thank the creative commons content producers for the Latex template of the thesis and the small picture of the measuring tape in Figure 1.4. Special thanks go to my sister for the designing the title page (twice).

The final thanks belong to my family, friends and significant other, for the support and all the love.

Declaration of authorship

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids. I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree. I have not submitted the doctoral thesis, or parts of it, to another academic institution and the thesis has not been accepted or rejected. Furthermore,I declare that no collaboration with commercial doctoral degree supervisors took place, and that the principles of Humboldt-Universität zu Berlin for ensuring good academic practice were abided by.

Berlin, May 10, 2017

Katja Tummler

A

Appendix

A.1 Normalization and correlation of ALF and AQUA datasets

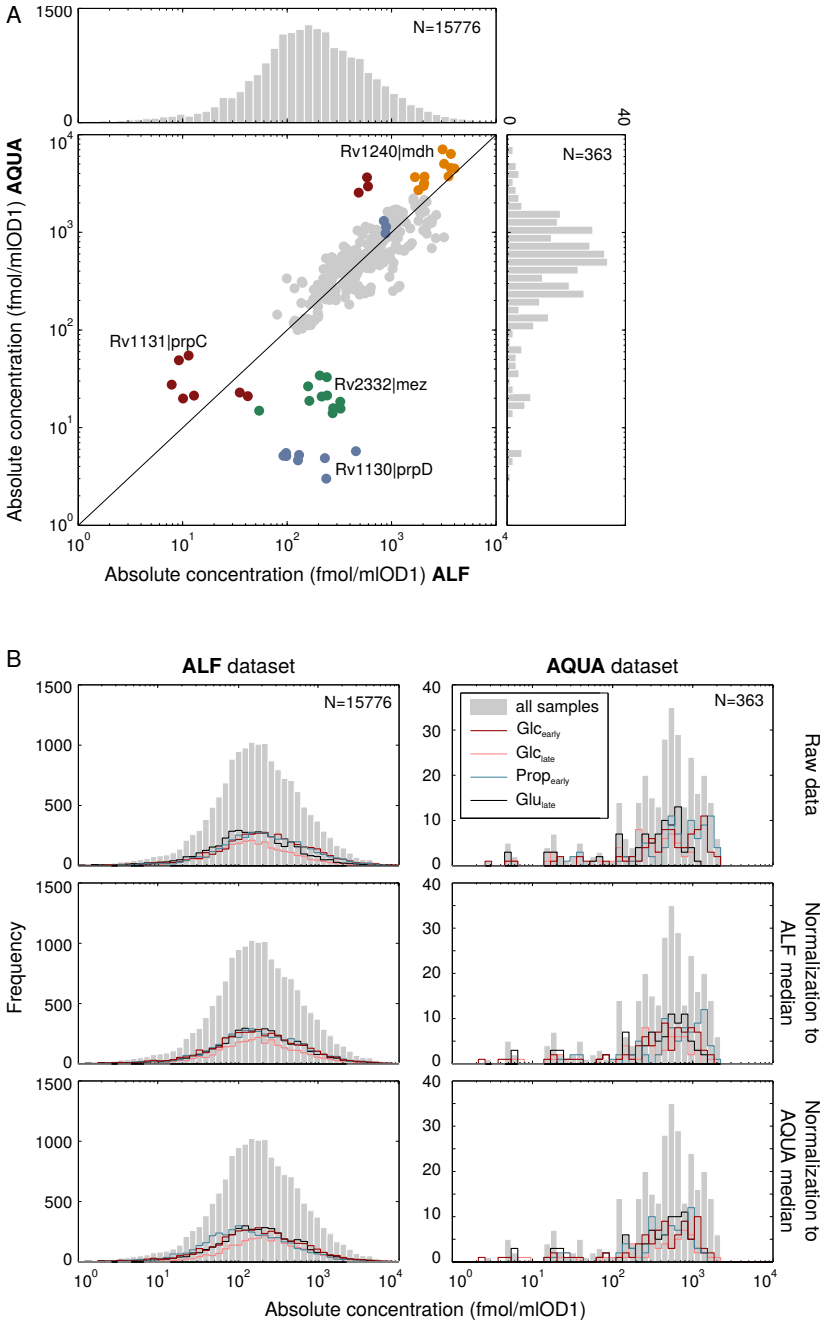


Figure A.1: **A** Correlation of the ALF and the Aqua datasets. Distributions for both data sets are plotted on the margins. Proteins strongly deviating between the two methods are colored and labeled with their gene names. Please note that the marginal distribution of the ALF dataset contains all measured proteins, whereas the scatter-plot only the ones that were found in both datasets. **B** Summary of the normalization of the protein data, showing the overall distribution (gray bars) and the distributions for the individual samples (colored lines). Upper panel shows the raw data with clearly deviating means, lower panels the data after normalization to the mean of either ALF or AQUA measurements. The normalization shown in the middle was applied for the subsequent use of the data.

A.2 BCG reference experiment

General bounds for growth and uptake rates were estimated based on data from batch growth of *M. bovis* BCG in media containing glycerol and glutamate (BCG reference experiment). The growth rate as well as the uptake and secretion rates of several media constituents were estimated using an exponential growth model and cell number dependent secretion or uptake. Only the measurements that could clearly be associated with exponential growth were considered. Errors for each component were estimated (*sd* as the % standard deviation, uniting experimental and biological variation). The results are summarized in Figure A.2 and Table A.2. The division time calculates as 37.9 h from $t_d = ln(2)/r_g$. For the default bound on uptake and secretion fluxes (Table 4.2), the below rates were normalized by the number of cells per mL at OD_{600}=1, the molecular weight of the compounds and the intracellular volume (*cf.* Table 4.1).

Table A.2: Estimated growth, uptake and secretion rates, initial concentrations and an estimator for the experimental error. For the media compounds, negative rates signify an uptake, positive ones secretion.

	unit	initial concentration	rate k [1/min]	error estimate *sd*
glycerol	g/L	3.55	$-9.93 \cdot 10^{-5}$	4
glutamate	g/L	0.37	$-9.82 \cdot 10^{-5}$	30
succinate	g/L	0.00133	$2.23 \cdot 10^{-13}$	17
aconitate	g/L	0.00005	$2.30 \cdot 10^{-6}$	23
cit_isocit	g/L	0.21	$9.39 \cdot 10^{-7}$	6
isocitrate	g/L	0.00	$2.81 \cdot 10^{-6}$	30
citrate	g/L	0.21	$-1.85 \cdot 10^{-6}$	6
OD_{600}	OD_{600}	0.12	0.00030437	13

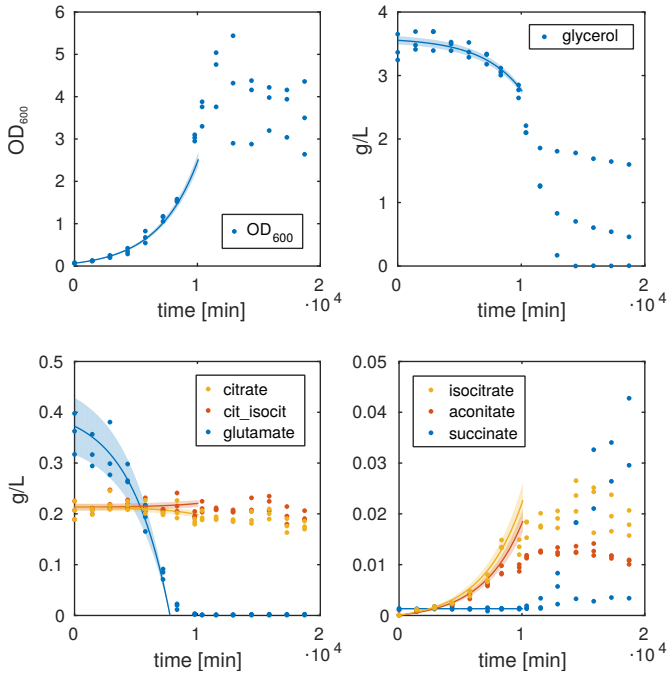


Figure A.2: BCG Reference experiment, data (dots) alongside model fit (lines) with error estimate (shaded area).

A.3 Growth rate dependency on ribosomal fraction

Organism	r_g [1/h]	t_d (min)	Ribosomal percentage f_R %(w/w)	translational capacity r_T (aa/s)	$f_R \cdot r_T$ (aa/min/ proteome)	Reference
E. coli	1.73	24	21.1	21	7.39	BNID 102345
	1.39	30	17.5	20	5.83	
	1.04	40	14.8	18	4.44	
	0.69	60	11.4	16	3.04	
	0.42	100	9.0	12	1.80	
M. bovis BCG	0.03	1200	3.0	2	0.10	[43], proteomics

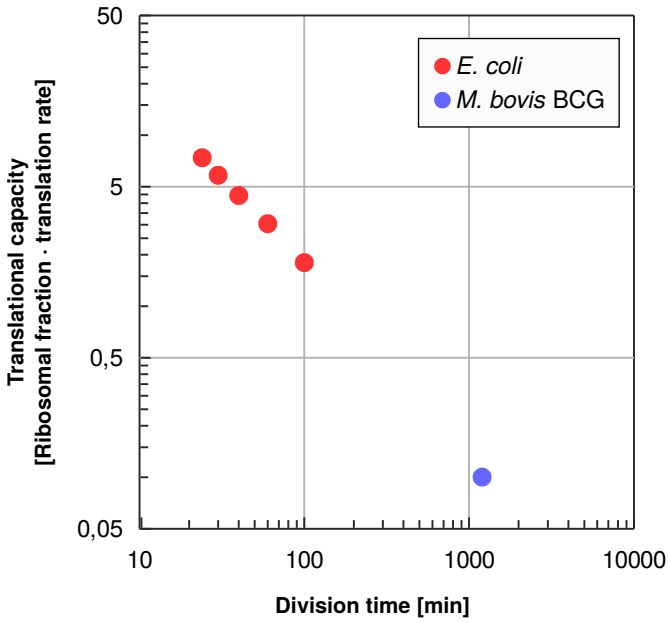


Figure A.3: Translational capacity as a function of the division time for *E. coli* cultures with different growth rates and for *M. bovis* BCG, based on the experimental data, explaining the slow growth rate of the mycobacteria.

A.4 GO enrichment of differentially expressed proteins

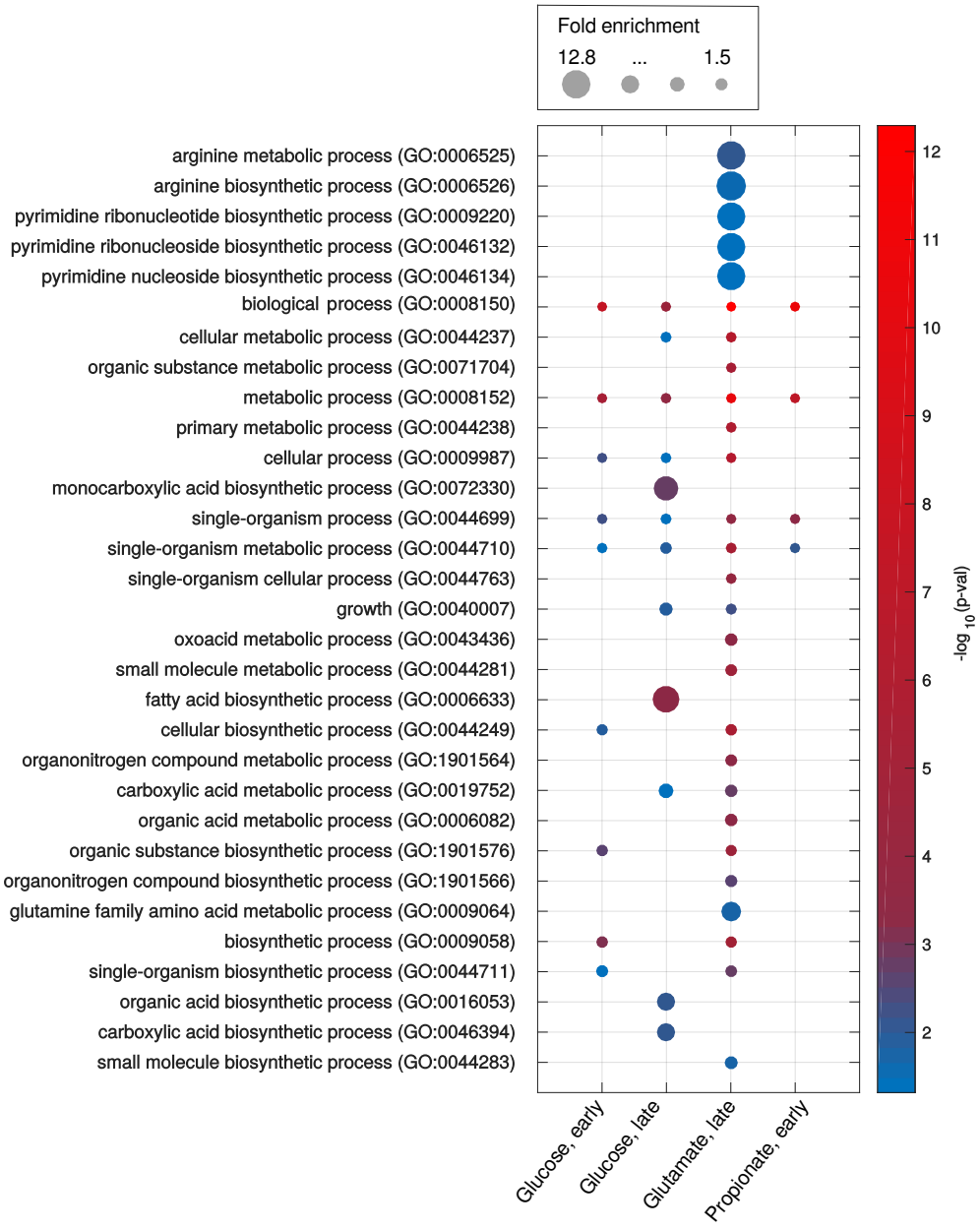


Figure A.4: Gene ontology terms enriched in the sets of differentially expressed genes for each of the four culture conditions. Significant terms are listed on the y-axis, larger dots indicate stronger enrichment, the color of the dots represents the statistical significance (red - low p-value, blue - higher p-value).

A.5 Principal component analysis metabolite concentrations

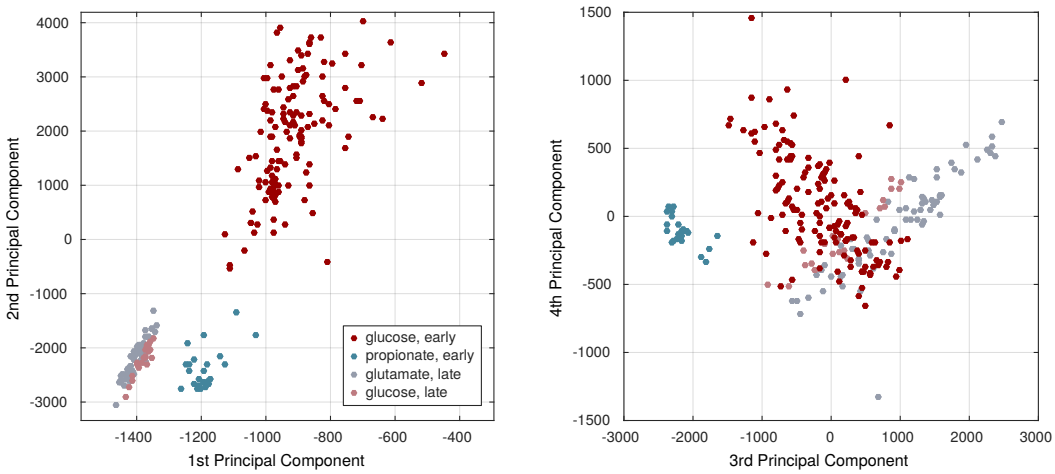


Figure A.5: Principal component analysis of the acquired metabolite data revealing similar culture conditions. Later growth phase samples (glucose - light red and glutamate - gray) cluster together. In the principal component analysis missing metabolite values were substituted with zeros. The first four principle components are shown, explaining 83.9, 12.6, 2.7 and 0.3 % of the total variance, respectively. The first component is shown in zoom in, maximum values reach up to $3.9 \cdot 10^4$.

A.6 Condition dependent metabolite concentrations - controls

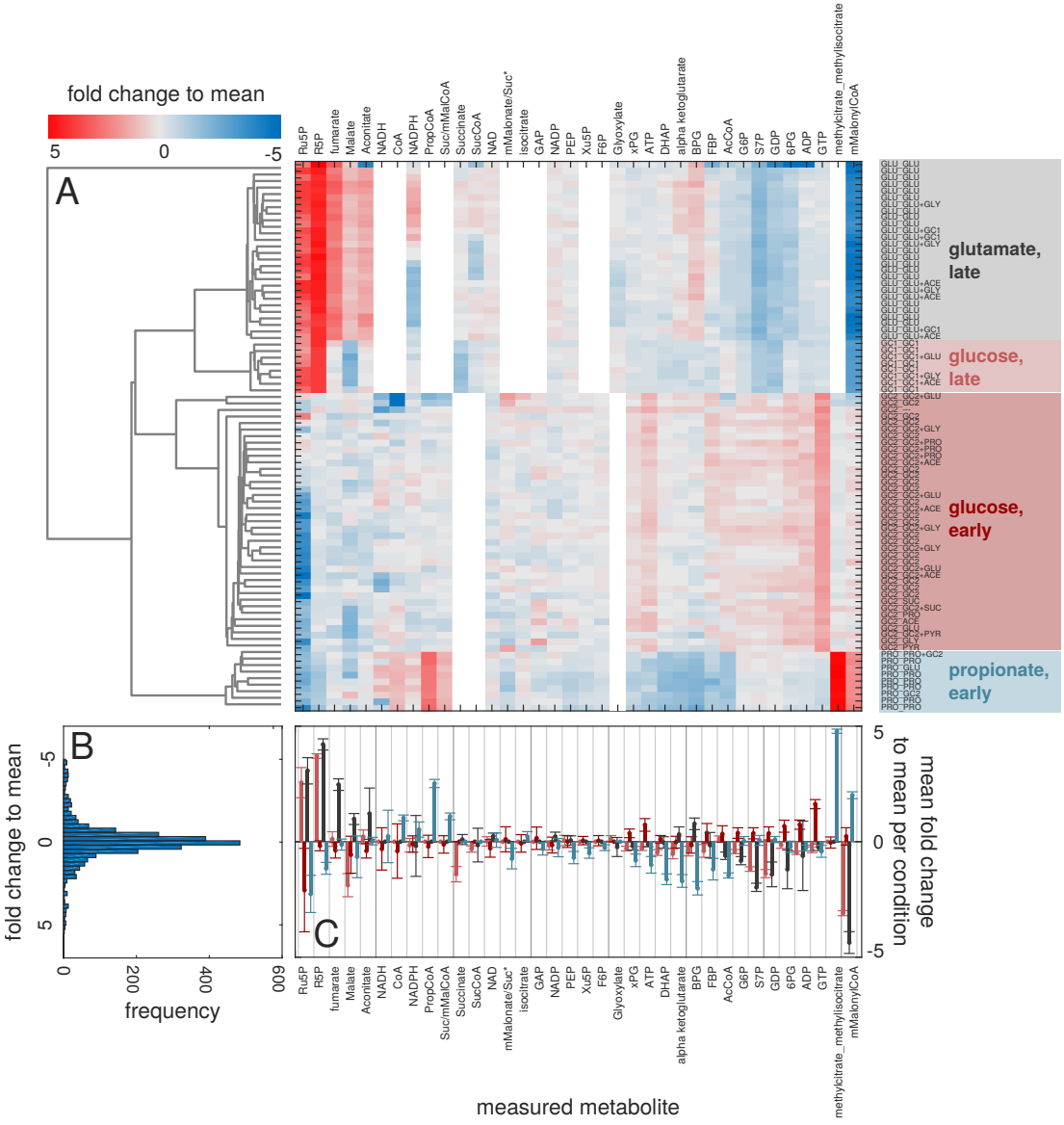


Figure A.6: A very busy figure on the steady state metabolite concentrations. **A** Clustered heatmap of all samples corresponding to a metabolic steady state (control experiments, samples at t=0), color represents the fold change to the mean of all shown samples. White regions indicate lacking data. Rows are labeled with sample names. **B** Histogram of all values in A. **C** Mean of the values in A for each condition and metabolite, colors correspond to the label groups in A. Errobars indicate the standard deviation.

A.7 Dynamic change in metabolite concentrations at t = 160s

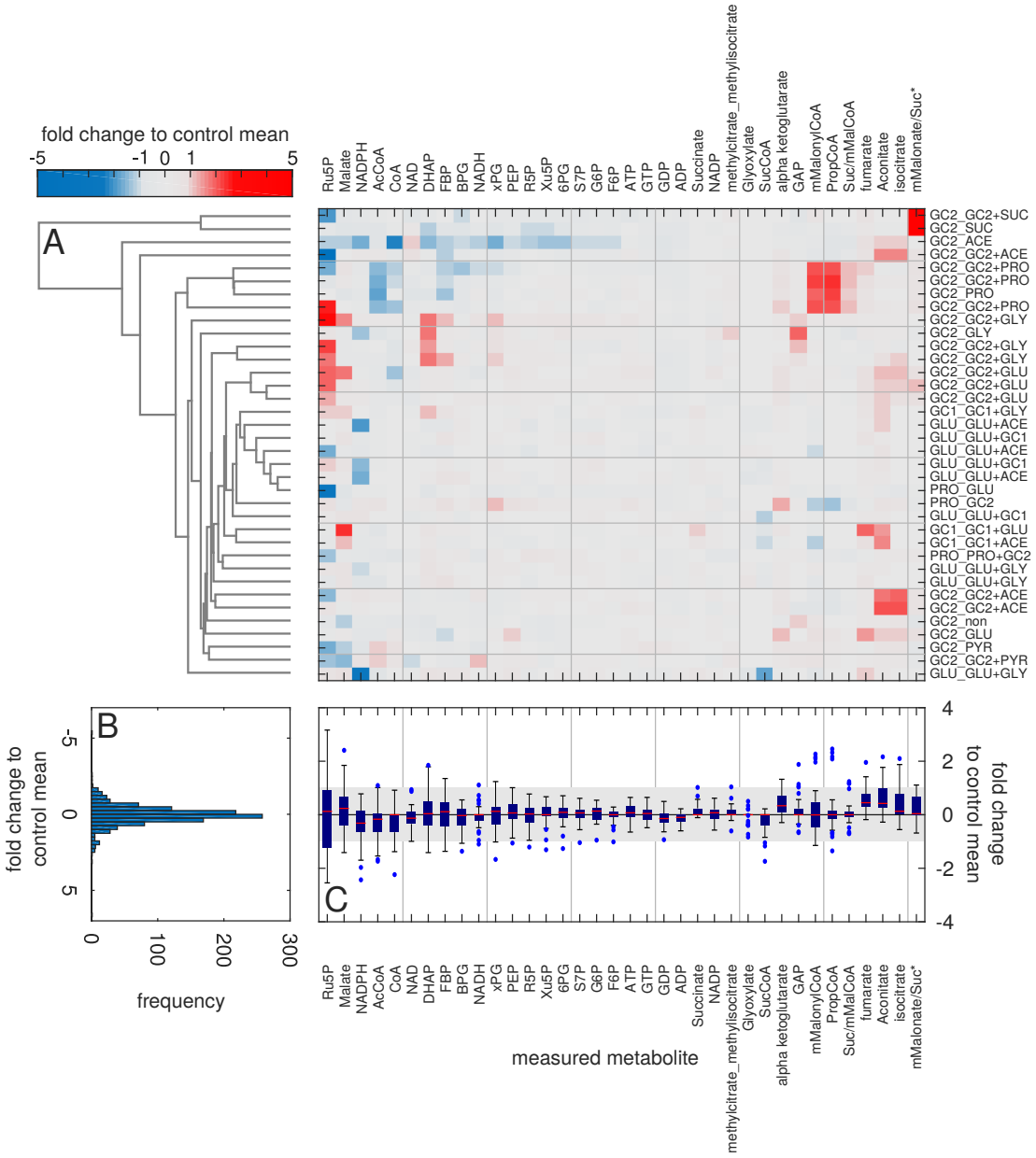


Figure A.7: Change in metabolite concentrations between the steady state and the perturbed state after 160s. **A** Clustered heatmap of all samples corresponding to the final measurement point after 160 s in the new media, color represents the $\log_2$ fold change to the mean of the respective steady state concentration of the metabolite in the specific culture condition (FC = $\log 2[\hat{y}_{ss}/\hat{y}(t = 160s)]$). Rows are labeled with sample names. Fold changes with an absolute value smaller than one are shown in gray. **B** Histogram of all values in A. **C** Boxplot of the values in A for each metabolite.

A.8 Dynamic change in metabolite concentrations - all timepoints

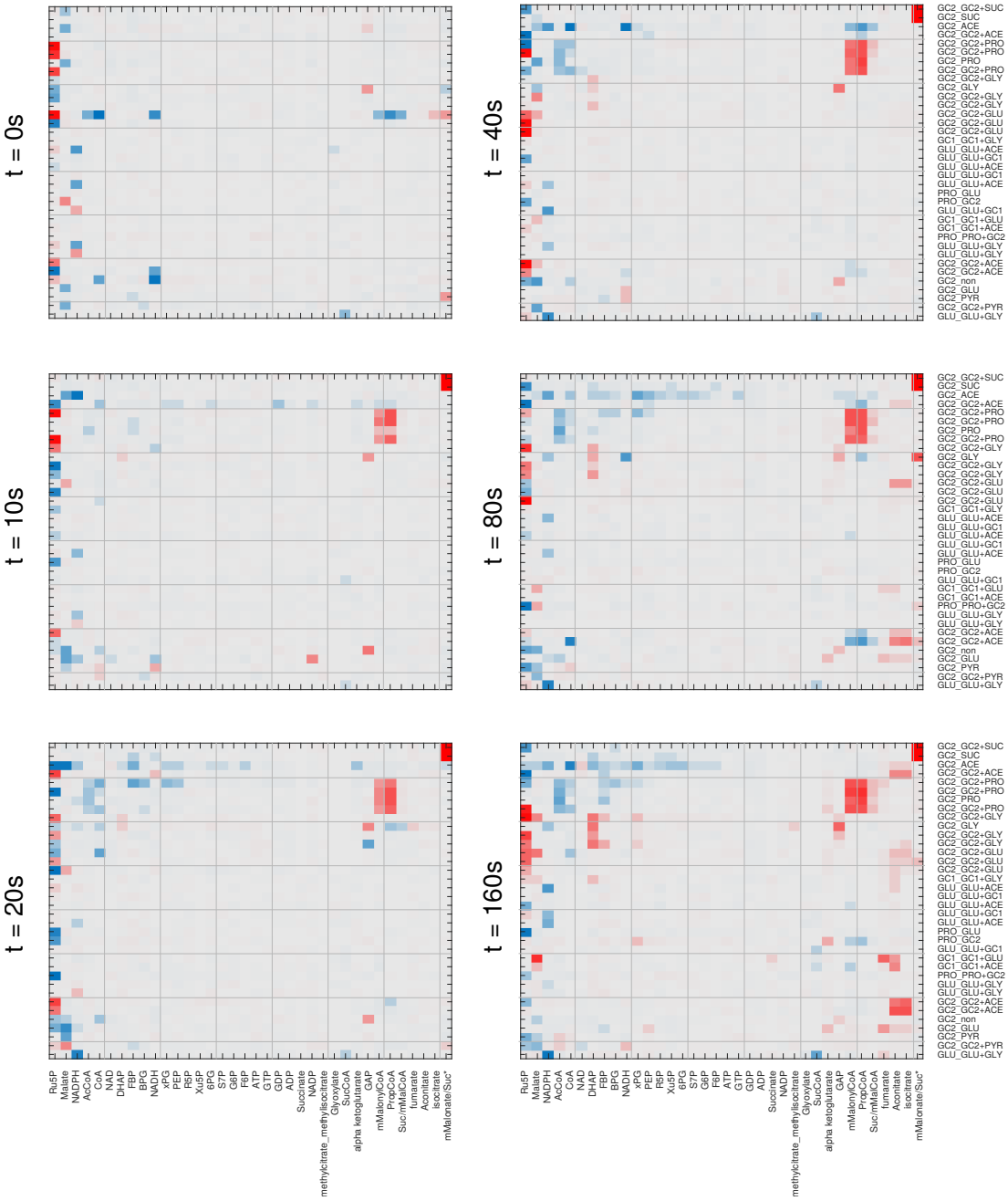


Figure A.8: Change in metabolite concentrations between the steady state and the dynamic measurements time points. Heatmaps are as in Appendix A.7 A, but for all timepoints as indicated.

A.9 List of modifications to sMTB

- The respiratory chain was amended according to the literature to guarantee its functionality:
  CYDB reaction did not pump protons, protons were added
  SDHA reaction was amended to charge one quinone under the consumption of two external protons (see Chapter 6.1)
- Net uptake of external protons Hr[e] was not allowed to avoid generation of energy without cost. As only Hr[e] are used in the ATPase-reactions, the import of the additional proton species H[e] does not need to be considered as it implements only an energy neutral release of excess protons (leakage).
- Excretion of carbon monoxide (CO[e]) was limited and nitrate respiration was prohibited, which was required for a meaningful flux distribution in glutamate conditions.
- The reactions PCA and PCKA were made reversible to allow for gluconeogenesis. They were only anaplerotic in sMtb.
- The presence of pyruvate dehydrogenase in *M. tuberculosis* is debated. It was added here to the sMtb model (as PYD), as its components are present in the proteomics data and based on literature evidence [194, 183, 13]

A.10 Biomass leaks from CCMtb

Table A.10: Leaks from CCMtb to produce biomass in mmol/gDW.

metKEGGID	Model Metabolite	Glucose	Propionate	Glutamate
C00001	H2O	9.71	25.52	-17.32
C00002	ATP	-10.04	-35.25	-35.26
C00003	NAD	10.23	8.18	-6.87
C00004	NADH	-10.23	-8.18	6.87
C00005	NADPH	-2.22	0	-13.23
C00006	NADP	2.22	0	13.23
C00008	ADP	10.04	35.25	35.26
C00009	PI	22.24	32.04	32.05
C00010	COA	0.73	1.44	1.43
C00011	CO2	6.74	5.66	5.66
C05819	MQH2	0	0	0
C00022	PYR	-7.36	-6.16	-6.16
C00024	ACCOA	-0.28	0	0
C00026	AKG	-0.62	-0.62	-0.62
C00031	GLC	-0.65	0	0
C00035	GDP	9.39	-6.67	-6.67
C00036	OA	-1.04	-1.04	-1.04
C00042	SUCC	0.18	0.19	0.18
C00044	GTP	-9.39	6.67	6.67
C00048	GLX	0	0	0
C00074	PEP	-0.36	-0.36	-0.36
C00085	F6P	-0.35	-1.00	-1.00
C00091	SUCCOA	-0.18	-0.19	-0.18
C00092	G6P	-0.48	-0.48	-0.48
C00100	PROPIONYLCOA	-0.26	-1.13	-1.13
C00111	DHAP	-0.21	-0.21	-0.21
C00117	R5P	-0.71	-0.71	-0.71
C00118	G3P	-1.14	-1.14	-1.14
C00122	FUM	0.46	0.46	0.46
C00149	MAL	0	0	0
C00158	CIT	0	0	0
C00197	3PG	-0.62	-0.62	-0.62
C00199	RL5P	-0.03	-0.03	-0.03
C00231	X5P	0	0	0
C00232	SUCCSAL	0	0	0
C00236	13PDG	0	0	0
C00279	E4P	-0.11	-0.11	-0.11
C00288	HCO3	0	0	0
C00311	ICIT	0	0	0
C00345	D6PGC	0	0	0
C00354	FDP	0	0	0
C00417	CISACONITATE	0	0	0
C00631	2PG	0	0	0
C00683	SMMALONYLCOA	-5	-0.12	-0.12
C01213	RMMALONYLCOA	0	0	0
C01236	D6PGL	0	0	0
C00828	MQ	0	0	0
C02225	2METHYLCITRATE	0	0	0
C04225	METHYLACONITATE	0	0	0
C04593	METHYLISOCITRATE	0	0	0
C05382	S7P	0	0	0

A.11 Leaks from CCMtb under propionate and glutamate conditions

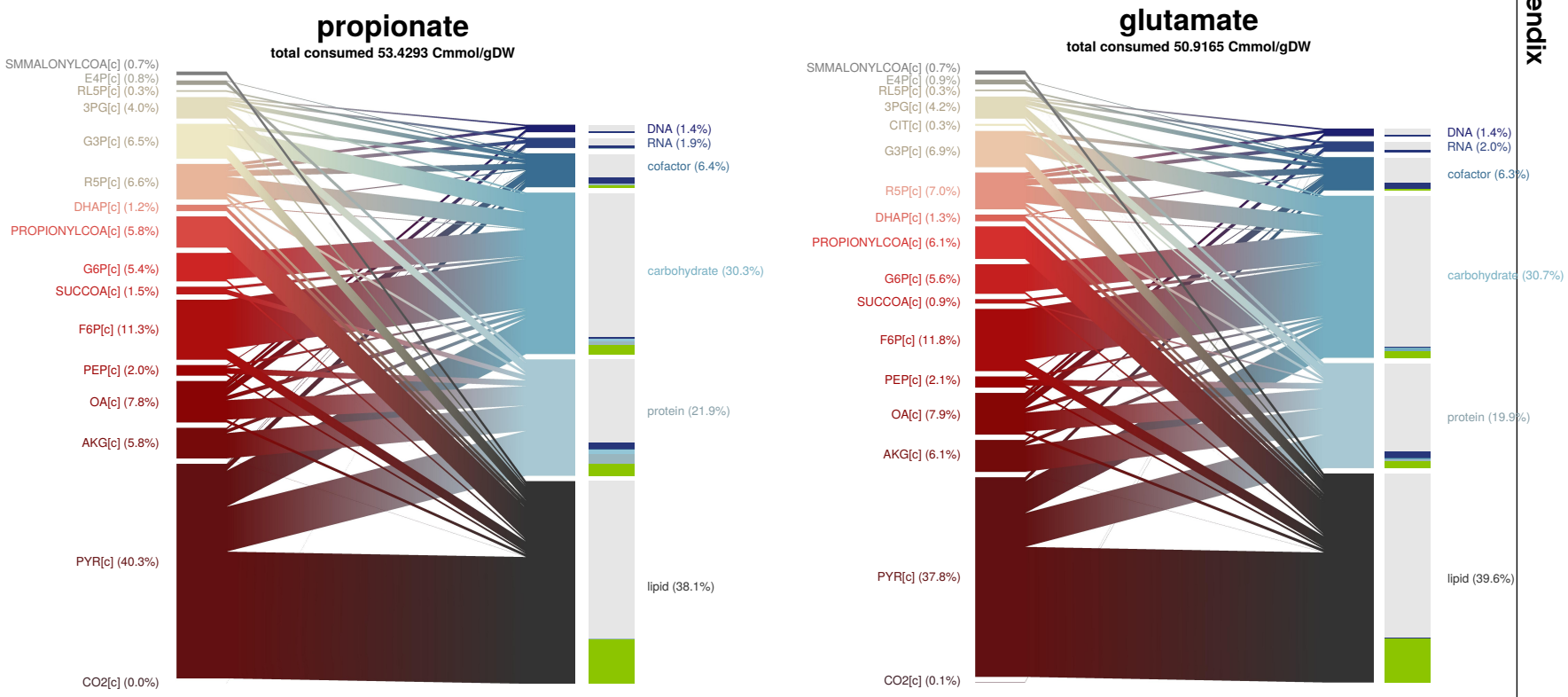


Figure A.11: Distribution of leaks from CCMtb intermediates (left side labels) to biomass constituents (right side labels) for propionate and glutamate conditions. For a more detailed explanation see Figure 5.4.

A.12 Released metabolites during the production of biomass

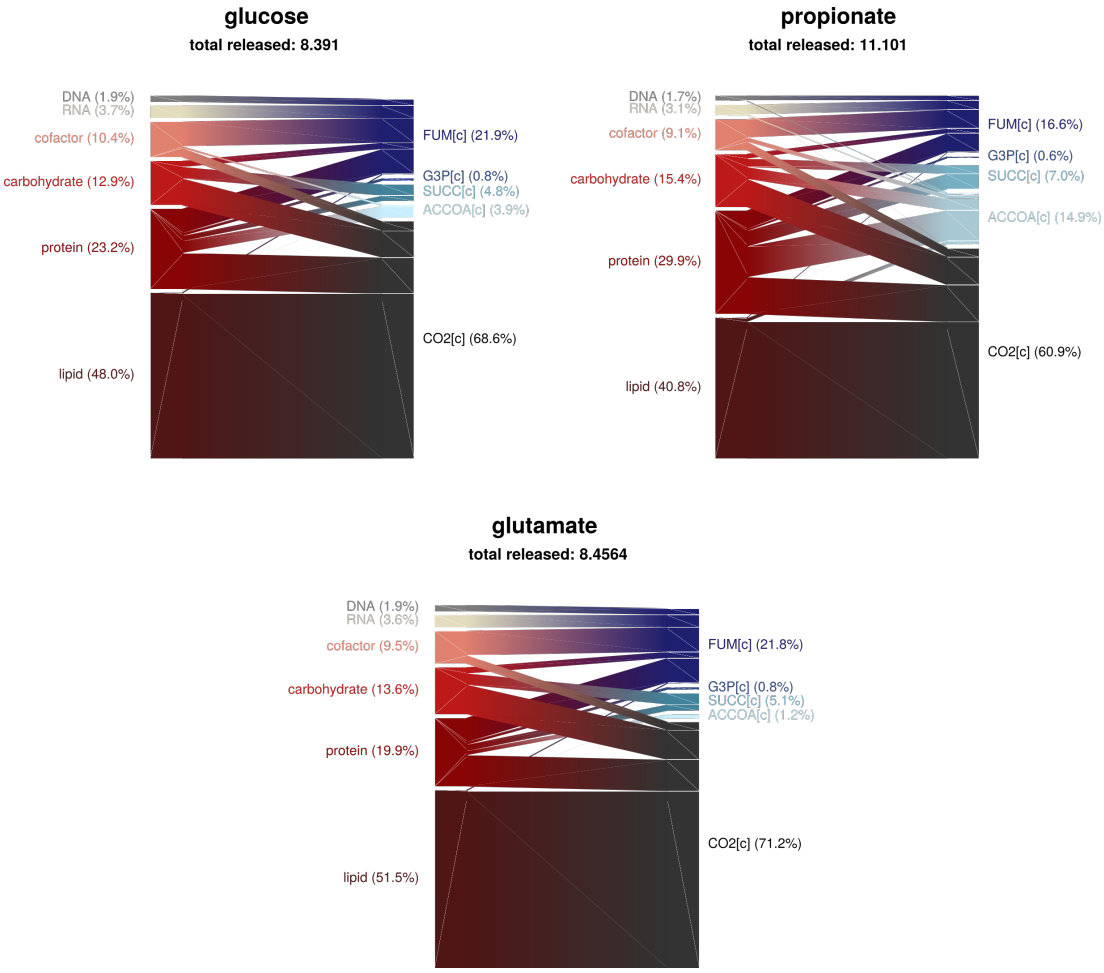


Figure A.12: Leaks of metabolites re-entering the central carbon metabolism from biosynthetic paths in mmol/gDW, corresponding to the stacked bars in Figures 5.4 and A.11. Mainly $CO_2$is released in decarboxylation reactions, larger amounts of fumarate (FUM[c]) get released during the synthesis of arginine and nucleotides via 5-Aminoimidazole-4-carboxamide ribonucleotide (AICAR).

A.13 Comparison between whole biomass and component backtracking

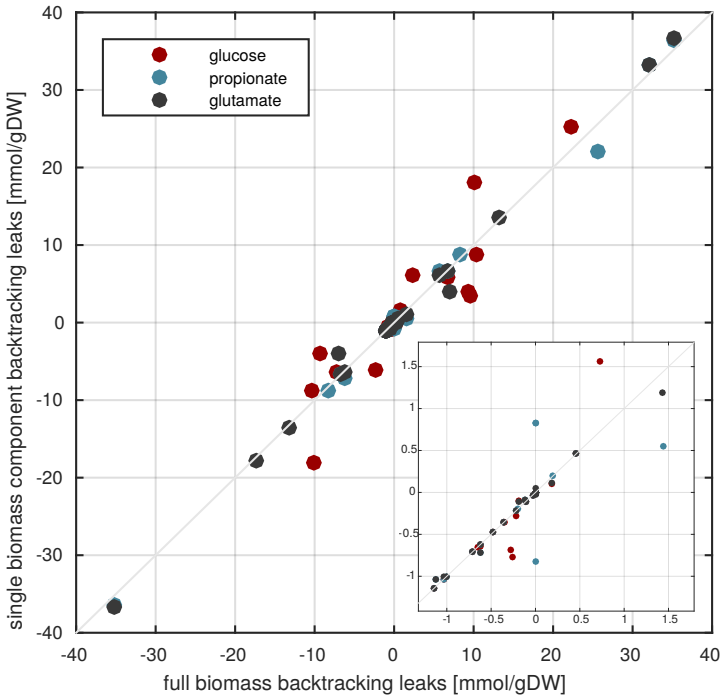


Figure A.13: Correlation of leak fluxes calculated by backtracking of the whole biomass (x-axis) and of the single biomass components (y-axis, corrsponding to the leaks shown in the distribution Figures 5.4 and A.11). Inset shows a close up of smaller leaks.

A.14 Loss of acquired carbon via different routes

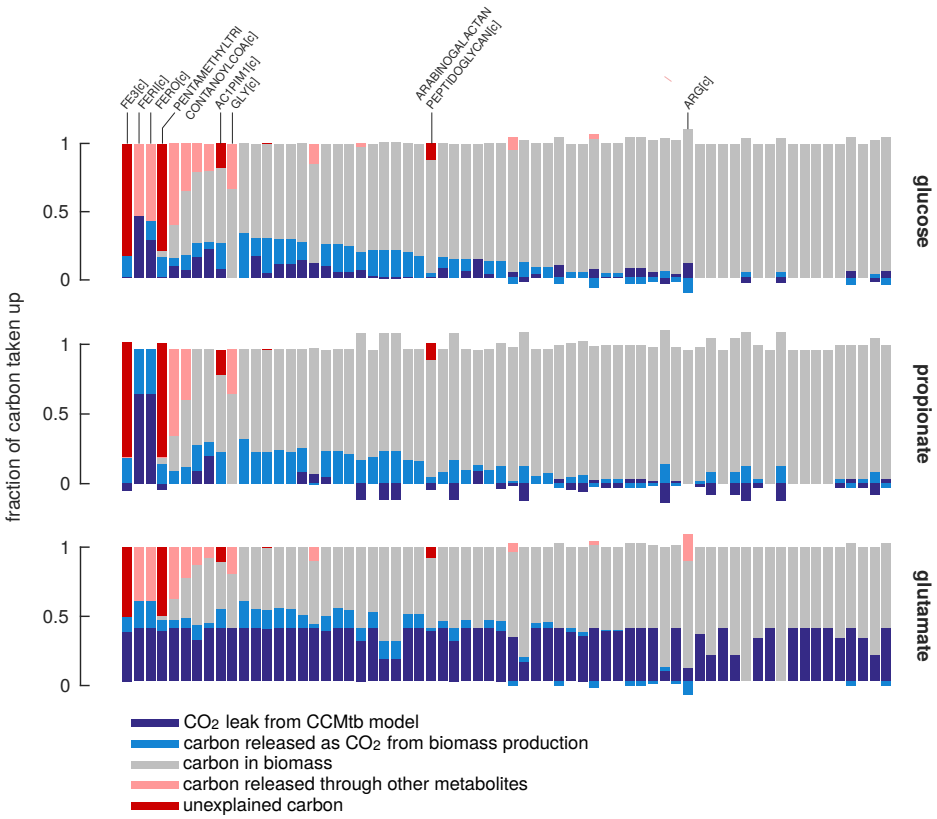


Figure A.14: Relative loss of acquired carbon via carbon dioxide and secretion of other metabolites in sMtb in the three growth conditions. For each biomass constituent, the gray bar represents the fraction of the carbon required from the central carbon metabolism that ends up in the biomass. Blue bars represent a release of carbon dioxide, either in the central carbon metabolism itself (dark blue) or along the remote biosynthetic paths (light blue). Light red bars imply loss of carbon due to the secretion of other metabolites into the medium and dark red bars represent unexplained carbon, hinting to an inconsistency in the genome scale model. Bars represent one biomass component each, are normalized to the number of carbons taken up to from one mole of each compound and ordered according to the fraction of carbon incorporated in the biomass in the glucose condition. Please note that the CO_2 leaks (blue) can be negative, representing reactions that fix CO_2 from the environment.

A.15 Michaelis-Menten constants of CCMtb reactions from literature

Table A.15: k^M values found in the literature for mycobacterial species. *Mtb- M. tuberculosis*, *Msmeg - Mycobacterium smegmatis*, *Mphl - Mycobacterium phlei*. * data measured by collaboration partners within the SystemTb Project (Davide Ferraris, Menico Rizzi, Università del Piemonte Orientale "A. Avogadro", Novara, Italy). Metabolite concentrations are as in Table A.18.

Reaction	EC number	Substrate	Value (mM)	Species	Reference
R02f_pgi	5.3.1.9	F6P	0.27	*Mtb*	[115]
R04f_fba	4.1.2.13	FDP	0.02	*Mtb*	[152]
R05f_tpi	5.3.1.1	G3P	84	*Mtb*	[116]
R05f_tpi	5.3.1.1	DHAP	0.0025	*Mtb*	[39]
R09_pyk	2.7.1.40	PEP	0.04	*Msmeg*	[85]
R09_pyk	2.7.1.40	ADP	0.23	*Msmeg*	[85]
R11f_cis	2.3.3.1, 2.3.3.3	OA	0.057	*Mtb*	*
R11f_cis	2.3.3.1, 2.3.3.3	ACCOA	0.16	*Mtb*	*
R13f_aco_2	4.2.1.3	ICIT	0.56	*Mtb*	[10]
R14f_idh1	1.1.1.42	ICIT	0.01	*Mtb*	[9]
R14f_idh1	1.1.1.42	NADP	0.125	*Mtb*	[9]
R17f_fum	4.2.1.2	FUM	2.63	*Mtb*	*
R18f_mdh	1.1.1.37, 1.1.1.299	NADH	0.0286	*Mphl*	[197]
R18f_mdh	1.1.1.37, 1.1.1.299	OA	0.0443	*Mphl*	[197]
R18f_mdh	1.1.1.37, 1.1.1.299	NAD	1.31	*Mtb*	*
R18f_mdh	1.1.1.37, 1.1.1.299	MAL	3.75	*Mtb*	*
R19f_pcx	6.4.1.1	PYR	0.3	*Msmeg*	[125]
R19f_pcx	6.4.1.1	ATP	1.26	*Msmeg*	[125]
R19f_pcx	6.4.1.1	HCO3	3	*Msmeg*	[125]
R20f_kgd	4.1.1.71	AKG	0.48	*Mtb*	[193]
R21f_gab	1.2.1.16, 1.2.1.79	SUCCSAL	0.0133	*Mtb*	[47]
R21f_gab	1.2.1.16, 1.2.1.79	NADP	0.0613	*Mtb*	[47]
R22af_icl1	4.1.3.1	GLX	0.14	*Mtb*	[124]
R22af_icl1	4.1.3.1	SUCC	0.412	*Mtb*	[124]
R22af_icl1	4.1.3.1	ICIT	0.045	*Mtb*	[124]
R23f_mls	2.3.3.9	ACCOA	0.01/0.03	*Mtb*	[185, 149]
R23f_mls	2.3.3.9	GLX	0.03/0.057	*Mtb*	[185, 149]
R28f_pck	4.1.1.32	GTP	0.017	*Msmeg*	[110]
R28f_pck	4.1.1.32	OA	0.168	*Msmeg*	[110]
R28f_pck	4.1.1.32	GDP	0.072	*Msmeg*	[110]
R28f_pck	4.1.1.32	PEP	0.213	*Msmeg*	[110]
R28f_pck	4.1.1.32	CO2	18.2	*Mtb*	[110]
R43_fbp	3.1.3.11	FDP	0.015	*Mtb*	[110]
R49f_rpi	5.3.1.6	R5P	1/1.8/3.7	*Mtb*	[162, 163, 164]
R50f_tkt1	2.2.1.1	R5P	0.8	*Mtb*	[60]
R52f_tkt2	2.2.1.1	F6P	0.6	*Mtb*	[60]
R52f_tkt2	2.2.1.1	X5P	0.4	*Mtb*	[60]
R54f_pcc	6.4.1.3	PROPIONYLCOA	0.24	*Mtb*	[62]

A.16 Comparison of different objective functions

Table A.16: Summary of the objective functions tested for resolving infeasibilites in the thermodynamic landscape of the central carbon metabolism. $\sum z$ denotes the number of infeasible reactions, $\sum w$ the total flux through all infeasible reactions, w_{pFBA} is the weight of the sum of fluxes in the overall objective, not converged - the algorithm did not find an optimum solution within 20 minutes of computation time.

Method		**Glucose**, early	**Glucose**, late	**Glutamate** late	**Propionate** early
MDF, w_{pFBA} =0.01	# unfeas. reactions	1 (tpi)	0	1 (pgk)	9 (fum, mdh, mez, pck, pcx, eno, gpm, pgk, gpd)
	runtime [s]	1.20	0.63	7.207938	0.91
	loops	no	no	no	no
$\sum z$	# unfeas. reactions	1 (tpi)	0	1 (pgk)	1 (pgk)
	runtime [s]	0.92	0.71	0.57	0.79
	loops	yes	yes	no	yes
$\sum z$, w_{pFBA} =0.001	# unfeas. reactions	1 (tpi)	0	9 (rpe, tkt, gpd, pgk, gpm, pcx, mdh, aco2, idh)	1 (pgk)
	runtime [s]	1.3	0.3	1.8	1.2
	loops	no	no	no	no
$\sum z$, MDF	# unfeas. reactions	0	-	1 (pgk)	1 (pgk)
	runtime [s]	21.95	not converged	2.620	1.71
	loops	yes	-	yes	yes
$\sum w$	# unfeas. reactions	0	-	-	-
	runtime [s]	1.5	not converged	not converged	not converged
	loops	yes	-	-	-
$\sum w$, w_{pFBA} =0.001	# unfeas. reactions	0	0	11 (scs, idh, aco2, aco1, cis, mdh, pcx, pgk, gpd, tkt1, rpe)	1 (pgk)
	runtime [s]	117.1	0.39	148.29	13.21
	loops	yes	no	no	no
$\sum w$, MDF	# unfeas. reactions	-	-	-	-
	runtime [s]	not converged	not converged	not converged	not converged
	loops	-	-	-	-

A.17 Optimal, thermodynamically feasible metabolite concentrations

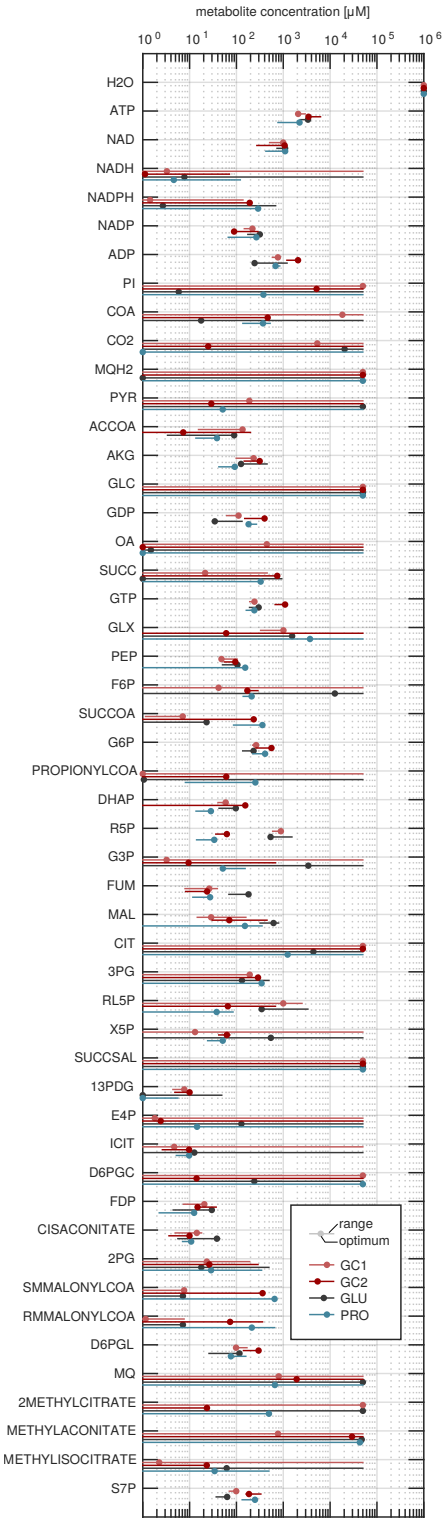


Figure A.17: Optimum metabolite concentrations (dots) for the thermodynamically feasible system for the four experimental conditions (dark red - glucose, early; light red - glucose late; gray - glutamate, late; blue - propionate, early) along with their allowed ranges (lines). Metabolites without experimental measurements were allowed to vary within the default concentration bounds (1 μM - 50 mM, see Table 4.2).

A.18 Summary of model metabolites

Table A.18: Metabolites included in the CCMtb model and Pdtx models. #C indicates the number of carbons.

ID	Name	Sum formula	ChEBI ID	KEGG ID	$\Delta_f G^0$	#C
H2O	Water	H_2O	15377	C00001	-156.1	0
ATP	Adenosine nucleotide triphosphate	$C_{10}H_{16}N_5O_{13}P_3$	15422	C00002	-2292.4	10
NAD	Nicotinamide adenine dinucleotide (oxidized)	$C_{21}H_{28}N_7O_{14}P_2$	15846	C00003	1053.6	21
NADH	Nicotinamide adenine dinucleotide (reduced)	$C_{21}H_{29}N_7O_{14}P_2$	16908	C00004	1115.0	21
NADPH	Nicotinamide adenine dinucleotide phosphate (oxidized)	$C_{21}H_{30}N_7O_{17}P_3$	16474	C00005	235.2	21
NADP	Nicotinamide adenine dinucleotide phosphate (reduced)	$C_{21}H_{29}N_7O_{17}P_3$	18009	C00006	173.0	21
ADP	Adenosine nucleotide diphosphate	$C_{10}H_{15}N_5O_{10}P_2$	16761	C00008	-1425.8	10
PI	Orthophosphate	H_3PO_4	18367	C00009	-1059.2	0
COA	Coenzyme A	$C_{21}H_{36}N_7O_{16}P_3S$	15346	C00010	1238.9	21
CO2	Carbon dioxide	CO_2	16526	C00011	-386.0	1
MQH2	Menaqinol	$C_{16}H_{18}O_2(C_5H_8)_n$	18151	C05819	-353.2	16
PYR	Pyruvate	$C_3H_4O_3$	32816	C00022	-351.2	3
ACCOA	Acetyl-CoA	$C_{23}H_{38}N_7O_{17}P_3S$	15351	C00024	1180.3	23
AKG	2-Oxoglutarate	$C_5H_6O_5$	30915	C00026	-633.7	5
GLC	D-Glucose	$C_6H_{12}O_6$	4167	C00031	-429.3	6
GDP	guanosine nucleotide diphosphate	$C_{10}H_{15}N_5O_{11}P_2$	17552	C00035	-248.3	10
OA	Oxaloacetate	$C_4H_4O_5$	30744	C00036	-714.6	4
SUCC	Succinate	$C_4H_6O_4$	15741	C00042	-530.7	4
GTP	guanosine nucleotide triphosphate	$C_{10}H_{16}N_5O_{14}P_3$	15996	C00044	-530.7	10
GLX	Glyoxylate	$C_2H_2O_3$	16891	C00048	-428.7	2
PEP	Phosphoenolpyruvate	$C_3H_5O_6P$	44897	C00074	-1188.6	3
F6P	D-Fructose 6-phosphate	$C_6H_{13}O_9P$	61553	C00085	-1317.4	6
SUCCOA	Succinyl-CoA	$C_{25}H_{40}N_7O_{19}P_3S$	15380	C00091	894.2	25
G6P	D-Glucose 6-phosphate	$C_6H_{13}O_9P$	4170	C00092	-1320.6	6
PROPIONYLCOA	Propanoyl-CoA	$C_{24}H_{40}N_7O_{17}P_3S$	15539	C00100	1270.7	24
DHAP	Glycerone phosphate	$C_3H_7O_6P$	16108	C00111	-1096.	3
R5P	D-Ribose 5-phosphate	$C_5H_{11}O_8P$	17797	C00117	-1233.7	5
G3P	D-Glyceraldehyde 3-phosphate	$C_3H_7O_6P$	29052	C00118	-1088.4	3
FUM	Fumarate	$C_4H_4O_4$	18012	C00122	-523.2	4
MAL	(S)-Malate	$C_4H_6O_5$	30797	C00149	-682.9	4
CIT	Citrate	$C_6H_8O_7$	30769	C00158	-965.6	6
3PG	3-Phospho-D-glycerate	$C_3H_7O_7P$	17794	C00197	-1347.3	3
RL5P	D-Ribulose 5-phosphate	$C_5H_{11}O_8P$	17363	C00199	-1233.3	5
X5P	D-Xylulose 5-phospate	$C_5H_{11}O_8P$	16332	C00231	-1106.7	5
SUCCSAL	Succinate semialdehyde	$C_4H_6O_3$	16265	C00232	-1106.7	4
13PDG	3-Phospho-D-glyceroyl phosphate	$C_3H_8O_{10}P_2$	16001	C00236	-2206.0	3
E4P	D-Erythrose 4-phosphate	$C_4H_9O_7P$	48153	C00279	-1316.5	4
ICIT	Isocitrate	$C_6H_8O_7$	30887	C00311	-959.0	6
D6PGC	6-Phospho-D-gluconate	$C_6H_{13}O_{10}P$	48928	C00345	-1565.0	6
FDP	D-Fructose 1.6-bisphosphate	$C_6H_{14}O_{12}P_2$	16905	C00354	-2205.9	6
CISACONITATE	cis-Aconitate	$C_6H_6O_6$	32805	C00417	-801.0	6
2PG	2-Phospho-D-glycerate	$C_3H_7O_7P$	17835	C00631	-1341.4	3
SMMALONYLCOA	(S)-Methylmalonyl-CoA	$C_{25}H_{40}N_7O_{19}P_3S$	15466	C00683	-1085.9	25
RMMALONYLCOA	(R)-Methylmalonyl-CoA	$C_{25}H_{40}N_7O_{19}P_3S$	15465	C01213	-812.0	25
D6PGL	D-Glucono-1.5-lactone 6-phosphate	$C_6H_{11}O_9P$	16938	C01236	-1383.8	6
MQ	Menaquinone	$C_{16}H_{16}O_2(C_5H_8)_n$	16374	C00828	-378.7	16
2METHYLCITRATE	2-Methylcitrate	$C_7H_{10}O_7$	30836	C02225	-1024.6	7
METHYLACONITATE	2-Methylcisaconitate	$C_7H_8O_6$	16717	C04225	-1313.2	7
METHYLISOCITRATE	(2S.3R)-3-Hydroxybutane-1.2.3-tricarboxylate;	$C_7H_{10}O_7$	15607	C04593	740.4	7
S7P	Sedoheptulose 7-phosphate	$C_7H_{15}O_{10}P$	15721	C05382	-979.1	7

A.19 Observables of the model

Table A.19: Summary of the available data and their mapping to the model entities (see Appendix A.18) via observables. The last four columns show mean steady state measurements for the experimental conditions in μM. Values without internal standard are shown in italics.

Observable name	Corresponding species in CCMtb	Glc, late	Glc, early	Glu, late	Pro, early
6PG	D6PGL	133.0 ± 15.2	219.2 ± 33.8	71.0 ± 19.6	117.4 ± 17.9
AcCoA	ACCOA	*75.3 ± 7.4*	*90.9 ± 13.4*	*46.4 ± 5.3*	*26.0 ± 1.5*
Aconitate	CISACONITATE	11.5 ± 2.9	6.8 ± 1.4	24.7 ± 8.3	8.9 ± 0.8
ADP	ADP	677.5 ± 39.7	1629.8 ± 189.0	724.4 ± 208.8	763.6 ± 34.5
alphaketoglutarate	AKG	165.9 ± 29.2	230.0 ± 36.2	287.9 ± 70.7	67.1 ± 10.9
ATP	ATP	2509.6 ± 186.8	4850.0 ± 604.8	2837.0 ± 233.7	1497.3 ± 315.9
BPG	13PDG	6.0 ± 0.7	7.4 ± 1.1	13.2 ± 2.6	1.7 ± 0.3
CoA	COA		*128.3 ± 41.8*		*330.6 ± 24.0*
DHAP	DHAP	49.7 ± 4.0	75.4 ± 5.7	69.7 ± 11.8	21.1 ± 3.2
F6P	F6P		229.5 ± 25.4		175.3 ± 15.6
FBP	FDP	14.0 ± 2.9	25.9 ± 4.8	17.1 ± 5.5	7.4 ± 2.2
fumarate	FUM	23.6 ± 6.8	15.9 ± 3.4	125.1 ± 24.4	19.6 ± 3.4
G6P	G6P	242.9 ± 8.8	411.3 ± 64.3	184.1 ± 20.9	316.0 ± 40.4
GAP	G3P		140.1 ± 39.1		102.2 ± 22.2
GDP	GDP	86.5 ± 10.6	274.3 ± 53.9	83.1 ± 21.1	224.7 ± 18.7
Glyoxylate	GLX	*668.1 ± 42.2*		*504.9 ± 129.1*	
GTP	GTP	217.3 ± 11.0	882.3 ± 91.9	244.3 ± 23.4	201.6 ± 17.2
isocitrate	ICIT		6.2 ± 1.6		7.5 ± 1.0
Malate	MAL	86.9 ± 31.4	240.1 ± 91.2	556.7 ± 103.2	164.3 ± 81.7
methylcitrate_methylisocitrate	2METHYLCITRATE + METHYLISOCITRATE		*12.2 ± 1.4*		*365.6 ± 15.9*
mMalonate/Suc*	SUCC		222.7 ± 64.6		104.2 ± 27.9
mMalonylCoA	SMMALONYLCOA + RMMALONYLCOA	*4.6 ± 0.4*	*51.0 ± 9.5*	*1.9 ± 0.7*	*160.8 ± 16.5*
NAD	NAD	754.1 ± 105.4	671.7 ± 173.3	914.4 ± 78.1	761.1 ± 146.0
NADH	NADH		*14.3 ± 7.0*		*18.8 ± 12.7*
NADP	NADP	183.5 ± 15.7	188.7 ± 43.3	245.3 ± 30.5	166.5 ± 43.3
NADPH	NADPH	*64.1 ± 9.2*	*73.7 ± 14.7*	*92.6 ± 73.8*	*97.0 ± 24.0*
PEP	PEP	75.6 ± 12.0	74.8 ± 8.6	77.7 ± 11.9	44.0 ± 8.0
PropCoA	PROPIONYLCOA		*23.1 ± 4.6*		*130.9 ± 15.3*
R5P	R5P	741.0 ± 63.9	49.3 ± 5.7	1033.8 ± 216.6	23.9 ± 4.2
Ru5P	RL5P	1208.4 ± 567.1	126.8 ± 241.4	1851.1 ± 654.6	42.2 ± 18.5
S7P	S7P	84.7 ± 6.2	247.1 ± 37.2	50.2 ± 5.7	190.8 ± 25.3
Suc/mMalCoA	SUCCOA + SMMALONYLCOA + RMMALONYLCOA		*100.7 ± 16.5*		*222.7 ± 16.9*
Succinate	SUCC	154.4 ± 37.1		391.8 ± 66.6	
SucCoA	SUCCOA	*4.1 ± 0.4*		*5.1 ± 2.2*	
xPG	3PG + 2PG	161.5 ± 13.3	246.3 ± 18.5	171.1 ± 23.4	103.7 ± 17.4
Xu5P	X5P		57.0 ± 6.6		37.6 ± 5.8

A.20 Enzyme complexes of CCMtb

Table A.20: Summary of enzyme complexes in CCMtb. * Values and information were extracted from the BRENDA database [175]. A complex stoichiometry of 1 was assumed for all unknown enzyme structures.

ID	Name	KEGGID	EC Number	RvNumber	Gene name	Comments	Ref.	Structure	Monomers/ active site
R01f_hex	glucokinase	R00299	2.7.1.2	Rv2702	ppgK	Poly-P or ATP can be used by Rv2702. Not sure if this is the correct reaction for 1stglycolytic step, also annotated with Rv0650, but much less information	[77]	unknown	1
R02f_pgi	glucose-6-phosphate isomerase	R00771	5.3.1.9	Rv0946c	pgi		[5]	dimer	2
R03_pfk	phosphofructokinase	R00756	2.7.1.11	Rv3010c	pfkA	Also Rv2029c (pfkB) but only pfkA detected in samples.	*	unknown, tetramer in many other species	4
R04f_fba	fructose-biphosphate aldolase	R01070	4.1.2.13	Rv0363c	fba		[144]	class II aldolase with 4 monomers	4
R05f_tpi	triosephosphate isomerase	R01015	5.3.1.1	Rv1438	tpiA		[207]	dimer with one active site	2
R06f_gpd	glyceraldehyde-3-phosphate dehydrogenase	R01061	1.2.1.12	Rv1436	gap		*	unknown, tetramer in many other species	4
R07f_pgk	phosphoglycerate kinase	R01512	2.7.2.3	Rv1437	pgk		*	unknown, monomer in many other species	1
R08f_eno	phosphopyruvate hydratase	R00658	4.2.1.11	Rv1023	eno		*	unknown, dimer in E.coli	2
R09_pyk	pyruvate kinase	R00200	2.7.1.40	Rv1617	pyk		*	unknown, tetramer in many other species	4
R10_pyd	pyruvate dehydrogenase	R00209	1.2.4.1, 1.8.1.4, 2.3.1.12	Rv2241	aceE	E1, was also Rv2496c AND Rv2497c (branched-chain alpha-keto acid dehydrogenase complex)			1
				Rv0462	lpdC	E3, also in kdh complex (was also Rv0794c AND Rv3303c)			1
				Rv2215	dlaT	E2, was also Rv2495c	[194]		1
R11f_cis	citrate synthase	R00351	2.3.3.1	Rv0889c	citA		[54]	dimer with one active site	2
			2.3.3.1	Rv0896	gltA2			unknown	1
R12f_aco_1	aconitate hydratase	R01325	4.2.1.3_a	Rv1475c	acn		[10]	monomer	1
R13f_aco_2	aconitate hydratase	R01900	4.2.1.3_b	Rv1475c	acn				1

ID	Name	KEGGID	EC Number	RvNumber	Gene name	Comments	Ref.	Structure	Monomers/ active site
R14f_idh1	isocitrate dehydrogenase	R00267	1.1.1.42	Rv3339c	icd	icd1	[150]	homodimer	2
				Rv0066c	icd2	icd2	[9]	homodimer	2
R15f_scs	succinyl-CoA synthase	R00405	6.2.1.5	Rv0952	sucD		*	in E. coli: heterodimer or 2*dimer	1
				Rv0951	sucC				1
R17f_fum	fumarate hydratase	R01082	4.2.1.2	Rv1098c	fumC		*	unknown, dimer or tetramer in many other species	2
R53_frd	fumarate reductase	R02164	1.3.5.4	Rv1552	frdA	flavoprotein subunit, frdA		unknown	1
				Rv1553	frdB	iron-sulfur subunit, frdB			1
				Rv1554	frdC	membrane anchor, frdC			1
				Rv1555	frdD	membrane anchor, frdD			1
R18f_mdh	malate dehydrogenase	R00342	1.1.1.37	Rv1240	mdh		[54]	dimer with one active site (?)	2
R19f_pcx	pyruvate carboxylase	R00344	6.4.1.1	Rv2967c	pca		[125]	tetramer in Msmeg,	4
R20f_kgd	2-oxoglutarate decarboxylase	R00272	4.1.1.71	Rv1248c	kgd		[203]	homodimer with one active site	2
			4.1.1.71	Rv0555	menD	menD has also been found to have this function, but is not annotated in KEGG			2
R20f_mez	malate oxidoreductase	R00214	1.1.1.38	Rv2332	mez	can also catalyze R00217	[54]	dimer	2
R21f_gab	succinate-semialdehyde dehydrogenase	R00714	1.2.1.16	Rv0234c	gabD1	also annotated in 1.2.1.79 (NADP)		unknown, tetramer in Ecoli	4
			1.2.1.16	Rv1731	gabD2	gabD2			4
R22af_icl1	isocitrate lyase	R00479	4.1.3.1	Rv0467	icl		[67]	tetramer (icl1)	4
			4.1.3.1	Rv1915	aceAa	icl2a		unknown, assume same as icl1	2
				Rv1916	aceAb	icl2b			2
R23f_mls	malate synthase	R00472	2.3.3.9	Rv1837c	glcB	similar to MSA in E.coli	[185]	Monomer	1
R24f_mcs	2-methylcitrate synthase	R00931	2.3.3.5	Rv1131	prpC	in KEGG as 2.3.3.1		dimer in E.coli	2
R25f_mcd	methylcitrate dehydratase	R04424	4.2.1.79	Rv1130	prpD		*	unknown, monomer in E.coli	1
R26f_mcl	isocitrate lyase (methyl)	R00409	4.1.3.30	Rv0467	icl	function carried out by icl1		tetramer	4

ID	Name	KEGGID	EC Number	RvNumber	Gene name	Comments	Ref.	Structure	Monomers/ active site
R28f_pck	phosphoenolpyruvate carboxykinase	R00431	4.1.1.32	Rv0211	pckA		*	unknown, monomer in Msmeg	1
R43_fbp	fructose 1,6-biphosphatase II	R04780	3.1.3.11	Rv1099c	glpX			unknown	1
R44f_zwf	glucose-6-phosphate dehydrogenase	R00835	1.1.1.49	Rv1447c	zwf2		*	unknown, dimer or tetramer in many other species	2
			1.1.1.49	Rv1121	zwf1				2
R45f_dev	6-phosphogluconolactonase	R02035	3.1.1.31	Rv1445c	pgl		*	unknown, monomer in some other species	1
R47f_gnd2	phosphogluconate dehydrogenase	R01528	1.1.1.44	Rv1122	gnd2		*	unknown, dimer in many other species	2
			1.1.1.44	Rv1844c	gnd1	not expressed in our samples			2
R48f_rpe	ribulose-phosphate 3-epimerase	R01529	5.1.3.1	Rv1408	rpe			unknown	1
R49f_rpi	ribose-5-phosphate isomerase	R01056	5.3.1.6	Rv2465c	rpiB		[162]	dimer	2
R50f_tkt1	transketolase	R01641	2.2.1.1	Rv1449c	tkt		[60]	dimer	2
R51f_tal	transaldolase	R08575	2.2.1.2	Rv1448c	tal			unknown	1
R52f_tkt2	transketolase	R01067	2.2.1.1	Rv1449c	tkt		[60]	dimer	2
R17r_sdh	succinate dehydrogenase	R00408	1.3.5.1	Rv0247c	Rv0247c	iron-sulfur subunit sdh1A	[41]	heterotertramer with one active site	1
				Rv0248c	sdhA	iron-sulfur subunit sdh1B			1
				Rv0249c	Rv0249c	membrane anchor, sdh1CD			2
			1.3.5.1	Rv3316	sdhC	membrane anchor, sdh2C	[41]	heterotertramer with one active site	1
				Rv3317	sdhD	membrane anchor, sdh2D			1
				Rv3318	sdhA	flavoprotein subunit, sdh2A			1
				Rv3319	sdhB	iron-sulfur subunit, sdh2B			1
R54f_pcc	propionyl-CoA carboxylase	R01859	6.4.1.3	Rv3285	accA3	Alpha-subunit, accA3	[62]		1
				Rv3280	accD5	Beta-subunit, accD5		two rings of trimers	1
				Rv3281	accE5	Eps-subunit AccE5		AccE5 increases activiy of AccA3/D5 (~3:1:1 ratio)	3
R55f_mcm	methylmalonyl-CoA mutase	R00833	5.4.99.2	Rv1492	mutA	mutA	[173]	heterodimer of mutA/B	1
				Rv1493	mutB	mutB			1

ID	Name	KEGGID	EC Number	RvNumber	Gene name	Comments	Ref.	Structure	Monomers/ active site
R58_macn	aconitate hydratase	R04425	4.2.1.99	Rv1475c	acn	in KEGG only as 4.2.1.3	[10]	monomer	1
R20f_kdh	2-oxoglutarate dehydrogenase complex	R08549	1.2.4.2,	Rv1248c	kgd	E1, kgd with other functions, 1.2.4.2 activity not annotated in Mtb	[203]	kgd is dimer	2
			1.8.1.4,	Rv0462	lpdC	E3, also in pdh complex, was also Rv0794c AND Rv3303c			1
			2.3.1.61	Rv2215	dlaT	E2			1
R56_gpm2	phosphoglycerate mutase	R01518	5.4.2.11	Rv0489	gpmA	was also Rv2228c AND Rv2419c AND Rv3214 AND Rv3837c (only expressed in Glu), and EC 5.4.2.12			2
				Rv3214	gpm2	gpm2/entD cofactor dependant (2,3 PG)	[205]	dimer	2
R57_rmcoa	methylmalonyl-CoA epimerase	R02765	5.1.99.1	Rv1322A	Rv1322A	Annotation unclear, not in KEGG, only expressed in early samples	[173]	unknown, dimer in other species	2

A.21 Reactions of CCMtb

Table A.21: Summary of the reactions of the central carbon metabolism as implemented in the models. Gibbs energies of reaction are given, along with their estimated standard deviation. Metabolite names are as in Appendix A.18.

ID	$\Delta_r G^0$ [kJ/mol]	Substrates		Products
R19f_pcx	-6.4 ± 1.5	ATP + CO2 + PYR	↔	ADP + PI + OA
R28f_pck	-13.4 ± 3.4	CO2 + GDP + PEP	↔	OA + GTP
R14f_idh1	6.1 ± 3.2	NADP + ICIT	↔	NADPH + CO2 + AKG
R20f_kgd	-15.5 ± 3.4	AKG	↔	CO2 + SUCCSAL
R18f_mdh	26.9 ± 0.3	NAD + MAL	↔	NADH + OA
R11f_cis	-38.4 ± 0.4	H2O + ACCOA + OA	↔	COA + CIT
R15f_scs	1.3 ± 1.3	ADP + PI + SUCCOA	↔	ATP + COA + SUCC
R53_frd	-14.5 ± 7.1	0.300000 MQH2 + FUM	→	SUCC + 0.300000 MQ
R17r_sdh	14.5 ± 7.1	SUCC + 0.300000 MQ	→	0.300000 MQH2 + FUM
R21f_gab	-39.4 ± 2.4	H2O + NADP + SUCCSAL	→	NADPH + SUCC
R17f_fum	-3.4 ± 0.3	H2O + FUM	↔	MAL
R12f_aco_1	8.3 ± 1.1	CIT	↔	H2O + CISACONITATE
R13f_aco_2	-0.7 ± 1.1	H2O + CISACONITATE	↔	ICIT
R20f_mez	13.0 ± 3.2	NAD + MAL	↔	NADH + CO2 + PYR
R43_fbp	-9.7 ± 0.7	H2O + FDP	→	PI + F6P
R09_pyk	-24.8 ± 0.4	ADP + PEP	↔	ATP + PYR
R10_pyd	-35.5 ± 3.3	NAD + COA + PYR	↔	NADH + CO2 + ACCOA
R03_pfk	-18.5 ± 0.7	ATP + F6P	→	ADP + FDP
R01f_hex	-19.5 ± 0.4	ATP + GLC	↔	ADP + G6P
R08f_eno	-4.1 ± 0.3	2PG	↔	H2O + PEP
R02f_pgi	2.5 ± 0.4	G6P	↔	F6P
R05f_tpi	5.5 ± 0.6	DHAP	↔	G3P
R06f_gpd	2.6 ± 0.4	NAD + PI + G3P	↔	NADH + 13PDG
R04f_fba	21.2 ± 0.5	FDP	↔	DHAP + G3P
R07f_pgk	-18.7 ± 0.4	ADP + 13PDG	↔	ATP + 3PG
R23f_mls	-36.2 ± 2.1	H2O + ACCOA + GLX	↔	COA + MAL
R22af_icl1	9.2 ± 1.4	ICIT	↔	SUCC + GLX
R55f_mcm	-7.6 ± 2.1	RMMALONYLCOA	↔	SUCCOA
R54f_pcc	-5.1 ± 10.4	ATP + CO2 + PROPIONYLCOA	↔	ADP + PI + SMMALONYLCOA
R26f_mcl	9.8 ± 3.0	METHYLISOCITRATE	↔	PYR + SUCC
R24f_mcs	-42.9 ± 3.4	H2O + OA + PROPIONYLCOA	↔	COA + 2METHYLCITRATE
R25f_mcd	-0.9 ± 5.6	2METHYLCITRATE	↔	H2O + METHYLACONITATE
R44f_zwf	-6.3 ± 1.3	NADP + G6P	↔	NADPH + D6PGL
R49f_rpi	-2.0 ± 0.8	RL5P	↔	R5P
R52f_tkt2	-10.1 ± 1.9	X5P + E4P	↔	F6P + G3P
R47f_gnd2	10.9 ± 3.2	NADP + D6PGC	↔	NADPH + CO2 + RL5P
R48f_rpe	-3.4 ± 1.2	RL5P	↔	X5P
R50f_tkt1	-3.9 ± 1.9	R5P + X5P	↔	G3P + S7P
R45f_dev	-24.6 ± 1.7	H2O + D6PGL	↔	D6PGC
R51f_tal	-0.7 ± 1.4	G3P + S7P	↔	F6P + E4P
R56_gpm2	4.1 ± 0.7	3PG	↔	2PG
R57_rmcoa	0 ±5.8	SMMALONYLCOA	↔	RMMALONYLCOA
R58_macn	0.9 ± 5.6	H2O + METHYLACONITATE	↔	METHYLISOCITRATE
R20f_kdh	-28.3 ± 7.7	NAD + COA + AKG	↔	NADH + CO2 + SUCCOA
RBIO_GC1	-	1204.5 ATP + 1227.2 NADH + 266.6 NADPH + 882.6 PYR + 33.5 ACCOA + 74.1 AKG + 77.8 GLC + 124.6 OA + 1126.4 GTP + 42.6 PEP + 42.3 F6P + 22.2 SUCCOA + 57.1 G6P + 31.4 PROPIONYLCOA + 25.6 DHAP + 85.1 R5P + 137.0 G3P + 75.0 3PG + 3.7 RL5P + 13.4 E4P + 0.06 SMMALONYLCOA	→	1165.0 H2O + 1227.2 NAD + 266.6 NADP + 1204.5 ADP + 2669.3 PI + 87.2 COA + 809.0 CO2 + 1126.4 GDP + 22.2 SUCC + 55.1 FUM

ID	$\Delta_r G^0$ [kJ/mol]	Substrates		Products
RBIO_GC2	-	1204.5 ATP + 1227.2 NADH + 266.6 NADPH + 882.6 PYR + 33.5 ACCOA + 74.1 AKG + 77.8 GLC + 124.6 OA + 1126.4 GTP + 42.6 PEP + 42.3 F6P + 22.2 SUCCOA + 57.1 G6P + 31.4 PROPIONYLCOA + 25.7 DHAP + 85.1 R5P + 137.0 G3P + 75.0 3PG + 3.7 RL5P + 13.4 E4P + 0.06 SMMALONYLCOA	→	1165.0 H2O + 1227.2 NAD + 266.6 NADP + 1204.5 ADP + 2669.3 PI + 87.2 COA + 809.0 CO2 + 1126.4 GDP + 22.2 SUCC + 55.1 FUM
RBIO_PRO	-	4229.7 ATP + 981.6 NADH + 739.4 PYR + 74.1 AKG + 800.6 GDP + 124.6 OA + 42.6 PEP + 120.1 F6P + 23.4 SUCCOA + 57.1 G6P + 135.4 PROPIONYLCOA + 25.7 DHAP + 85.1 R5P + 137.0 G3P + 75.0 3PG + 3.7 RL5P + 13.4 E4P + 13.9 SMMALONYLCOA	→	3062.6 H2O + 981.6 NAD + 4229.7 ADP + 3845.3 PI + 172.6 COA + 679.7 CO2 + 23.4 SUCC + 800.6 GTP + 55.1 FUM
RBIO_GLU	-	2078.6 H2O + 4230.9 ATP + 824.7 NAD + 1588.0 NADPH + 739.4 PYR + 74.1 AKG + 800.6 GDP + 124.6 OA + 42.6 PEP + 120.0 F6P + 22.2 SUCCOA + 57.1 G6P + 135.4 PROPIONYLCOA + 25.7 DHAP + 85.1 R5P + 137.0 G3P + 75.0 3PG + 3.7 RL5P + 13.4 E4P + 13.9 SMMALONYLCOA	→	824.7 NADH + 1588.0 NADP + 4230.7 ADP + 3846.5 PI + 171.4 COA + 679.4 CO2 + 22.2 SUCC + 800.6 GTP + 55.1 FUM
RUP_GC0	-		→	GLC
RUP_GLU	-		→	AKG
RUP_PRO	-	COA	→	PROPIONYLCOA
RUP_ACE	-		→	ACCOA
RUP_GLY	-		→	G3P
RUP_PYR	-		→	PYR
RUP_SUC	-		→	SUCC
RSEC_SUC	-	SUCC	→	
RSEC_ACO	-	ACCOA	→	COA
RSEC_ICT	-	ICIT	→	
RSEC_P	-	PI	↔	
RSEC_CO2	-	CO2	↔	
RSEC_H2O	-	H2O	↔	
Rresp_nadh	-45.2 ± 5.9	NADH + 2.3 MQ	↔	NAD + 2.3 MQH2
Rresp_atps	-306.2 ± 13.4	H2O + ADP + PI + MQH2	↔	ATP + MQ
Rresp_trd	-0.4 ± 0.7	NAD + NADPH	↔	NADH + NADP
Rresp_ndk	2.5 ± 2.6	ADP + GTP	↔	ATP + GDP

A.22 Used kinetic laws

Table A.22: Kinetic laws and for which reactions they were used in the dynamic models.

#	Rate law	Formula	Used in
1	Convenience kinetics	$v = E \cdot \frac{k_+^{cat} \cdot \prod_i (S_i/k_{s,i}^M) \; - \; k_-^{cat} \cdot \prod_j (P_j/k_{p,j}^M)}{\prod_i (1+S_i/k_{s,i}^M) + \prod_j (1+P_j/k_{p,j}^M) - 1}$	enzymatic reactions (CCMtb)
2	Thermodynamic Michaelis-Menten	$v = E \cdot k_+^{cat} \cdot \left(\frac{\prod_i (S_i/k_{s,i}^M)}{\prod_i (S_i/k_{s,i}^M) + \prod_j (P_j/k_{p,j}^M) + 1} \right) \cdot \left(1 - e^{\Delta_r G'/RT}\right)$	enzymatic reactions (tCCMtb, Pdtx)
3	Thermodynamic Michaelis-Menten without enzyme	$v = v_+^{max} \cdot \left(\frac{\prod_i (S_i/k_{s,i}^M)}{\prod_i (S_i/k_{s,i}^M) + \prod_j (P_j/k_{p,j}^M) + 1} \right) \cdot \left(1 - e^{\Delta_r G'/RT}\right)$	respiration and cofactor-converting reactions (CCMtb, tCCMtb)
4	Michaelis Menten	$v = k^{cat} \cdot \frac{S}{k^M + S}$	Secretion reactions (all), individual biomass reactions (tCCMtb, Pdtx)
5	Fixed flux	$v = k$	uptake reactions (all), biomass reaction (CCMtb)

A.23 Collection of known regulatory points in Mtb

Table A.23: Summary of known (potential) regulatory points in mycobacteria. Phosphorylation data for *M. tuberculosis* is taken from Prisic *et al.* [148] with the following experimental conditions: L - log phase growth on glucose, S - stationary phase on glucose, N - nitric oxide stress, C - peroxide stress, H - hypoxia, A - acetate as carbon source. Phosphorylation data for *M. bovis* is taken from Zheng *et al.* [215] with the follwing experimental conditions: *M. bovis* var BCG NCTC 5692 was grown in 5 L of Sauton's liquid medium for three weeks at 37 °C under stationary conditions. Phosphorylations on specific amino acid residues are listed (S,T,Y for serine, thyrosine and threonine, respectively) * As activators and inhibitors only compounds which are not educts of the reaction are listed, all entries were extracted from the Brenda database except for ** [203] and *** [51].

Reaction ID	Gene Mtb	Uniprot ID	Gene names	Gene BCG	Phospho Mtb	Phospho BCG	Activator*	Inhibitor *
R04f_fba	Rv0363c	P9WQA3	fba	BCG0401c	L	S18		
R05f_tpi	Rv1438	P9WG43	tpiA	BCG1499		S13; T17		
R09_pyk	Rv1617	P9WKE5	pyk	BCG1655			G6P	
R11f_cis	Rv0896	P9WPD5	gltA2	BCG0948	L,A,S			
R14f_idh1	Rv3339c	P9WKL1	icd	BCG3409c		S6		GLX, OA
	Rv0066c	O53611	icd2	BCG0097c	L			
R17f_fum	Rv1098c	P9WN93	fumC	BCG1158c		S14; Y18		
R18f_mdh	Rv1240	P9WK13	mdh	BCG1300	L			
R19f_pcx	Rv2967c	I6YEU0	pca	BCG2988c		T14	ACO	
R20f_kgd	Rv1248c	P9WIS5	kgd	BCG1308c	L		ACO**	
R20f_kdh	Rv1248c	P9WIS5	kgd	BCG1308c	L			
R20f_mez	Rv2332	P9WK25	mez	BCG2354	A			
R22af_icl1	Rv0467	P9WKK7	icl	BCG0507				MAL
R23f_mls	Rv1837c	P9WK17	glcB	BCG1872c				3PG,D6PGL, PEP
R28f_pck	Rv0211	P9WIH3	pckG	BCG0248	A			AKG
R43_fbp	Rv1099c	P9WN21	glpX	BCG1159c				MCIT***
R45f_dev	Rv1445c	P9WQP5	pgl	BCG1506c		S9		
R50f_tkt1	Rv1449c	P9WG25	tkt	BCG1510c		S2; T12		
R52f_tkt2	Rv1449c	P9WG25	tkt	BCG1510c		S2; T12		
R55f_mcm	Rv1492	P9WJK7	mutA	BCG1555		S10		
R56_gpm2	Rv3214	Q6MWZ7	gpm2	BCG3241		N-Term(Acetyl); T10; Y18		

A.24 Common regulatory edges in other bacterial species

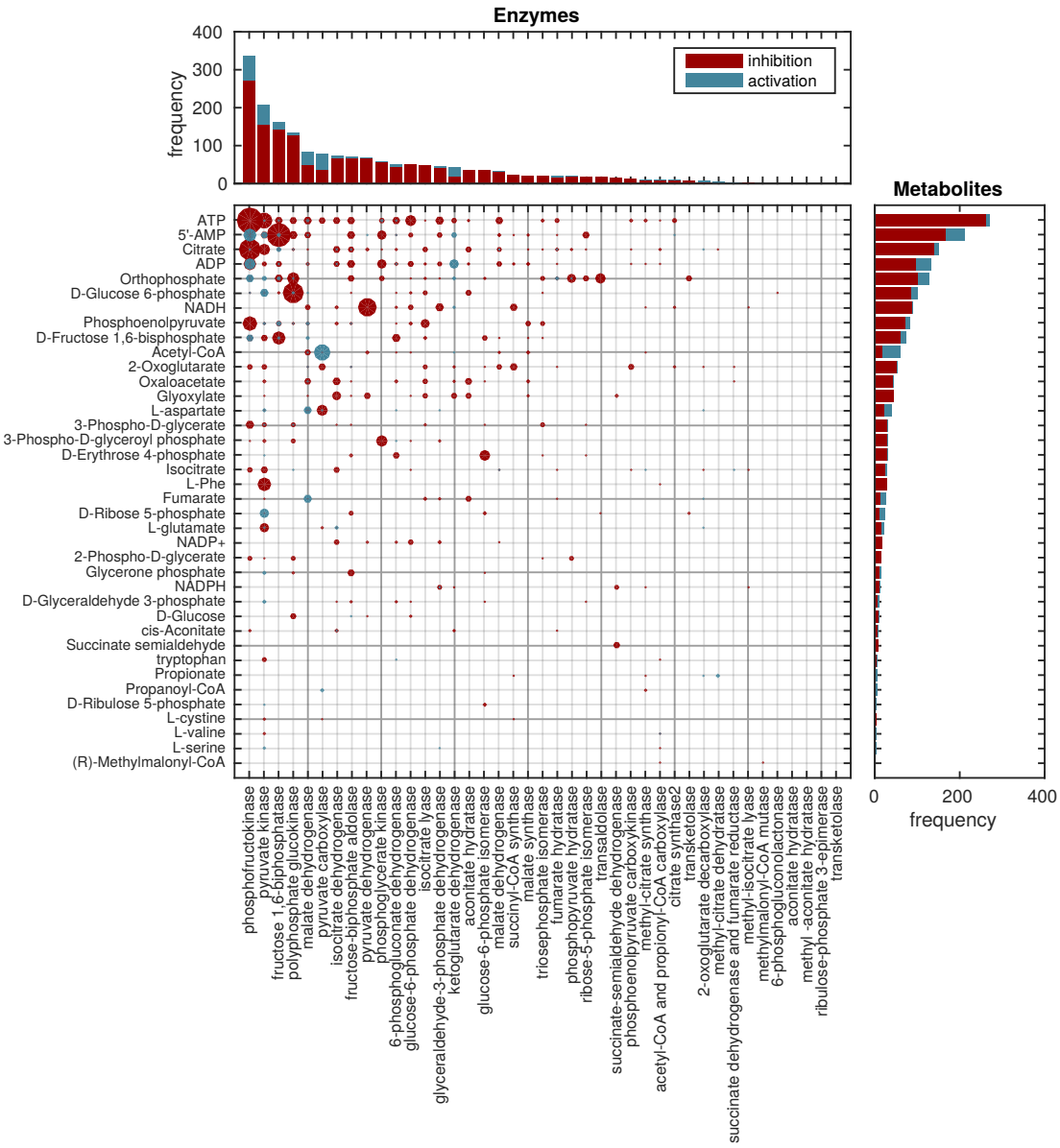


Figure A.24: Enzyme inhibitions and activations of the central carbon metabolic enzymes in mycobacterial species. Enzymes are shown on the x-axis, relevant metabolites on the y-axis. red dots indicate inhibitions, blue dots activations, the size of the dots scales with the number of found regulations. Stacked histograms of the total number of regulations is plotted on the margins. Data were obtained from the Brenda database.

A.25 Parameter histograms of single steady state fits

Figure A.25: Kernel density plots of the optimum parameters sets of ∼500 individual steady state fits for the three different nutrients. Each subplot extends from the lower to the upper bound set for the optimization. Non-overlapping distributions can hint to the need for regulation.

A.26 Maximum-minimum driving force optimization problem

The MDF calculations were carried out as described in [134] solving the optimization problem

$$\begin{aligned} \max_{\mathbf{x},\mathrm{MDF}} \quad & \mathrm{MDF} \\ \text{subject to} \quad & -(\mathbf{G^0} + R \cdot T \cdot \mathbf{N}^{\mathrm{T}} \cdot \ln \mathbf{x}) \geq \mathrm{MDF} \\ & \mathbf{x}_{lb} \leq \mathbf{x} \leq \mathbf{x}_{ub} \\ & \mathbf{y}_{lb} \leq \mathbf{O} \cdot \mathbf{x} \leq \mathbf{y}_{ub} \\ & \mathbf{\Delta}_r \mathbf{G}^0_{lb} \leq \Delta_r G^0 \leq \mathbf{\Delta}_r \mathbf{G}^0_{ub} \end{aligned}$$

with MDF as a tight lower bound equal to the Maximum-Minimum driving force in kJ/mol as defined by Noor et al. This value can is the negative of the specific Gibbs energy $\Delta_r G'$ of the reaction(s) in the pathway with the most shallow thermodynamic gradient (i.e. the lowest driving force). R and T denote the ideal gas constant in kJ/mol/K and the absolute temperature in K, respectively. The general boundaries on the model metabolite's concentrations $\mathbf{x}_{lb}, \mathbf{x}_{ub} \in \mathbb{R}^m$ contain the feasible logarithmic concentration ranges based on physiologically sensual values [57], in accordance with the ranges of the measured concentrations, between x_{min} = 1μM and x_{max} = 5·10^4μM. For measured metabolites the concentration vector $\mathbf{x}$ multiplied by the observables matrix $\mathbf{O}$ was confined to lie with in $\pm 2\sigma$ of the measured values $\mathbf{y} \in \mathbb{R}^y$ (collected in $\mathbf{y}_{lb}, \mathbf{y}_{ub} \in \mathbb{R}^y$). Note this additional constraint would be non linear in the logarithmic concentration space and we hence have to solve a nonlinear problem as opposed to the original linear MDF problem in [134].

A.27 Protein cost optimization problem

The optimization of the protein cost is based on the method in [57]

$$\begin{aligned} \max_{\mathbf{E},\mathbf{x}} \quad & \sum_{j=1}^{r} \lambda_j / V_{tot} \\ \text{where} \quad & \lambda_j / V_{tot} = 1/(Mw_j \cdot k^{cat}_{+,j} \cdot (\text{saturation term})_j \cdot (\text{thermodynamic term})_j) \\ \text{subject to} \quad & \mathbf{x}_{lb} \leq \mathbf{x} \leq \mathbf{x}_{ub} \\ & \mathbf{y}_{lb} \leq \mathbf{O} \cdot \mathbf{x} \leq \mathbf{y}_{ub} \\ & \mathbf{\Delta}_r \mathbf{G}^0_{lb} \leq \Delta_r G^0 \leq \mathbf{\Delta}_r \mathbf{G}^0_{ub} \end{aligned}$$

The saturation and kinetic terms are defined in the kinetic equation 7.1. Mw denotes the molecular weight of each enzyme complex in Da. Each λ_j corresponds to an entry in $\mathbf{E}$, describing the enzyme amount for reaction j in mol. V_{tot} is set to 1 mol/s to normalize the protein to a unit of flux, such that the calculated enzyme abundance as depicted in Figure 7.3 has the unit mol/(mol/s).

A.28 Flux maximization with limited enzyme allowance problem

For the maximization of the pathway flux given a limited amount of total enzyme E_{tot} in Da we solved the following optimization problem for values of f between 0 and 1:

$$\begin{aligned} \max_{\mathbf{E},\mathbf{x}} \quad & \sum_{i=P_1}^{P_2} v_{i,1} \\ \text{subject to} \quad & \mathbf{x}_{lb} \leq \mathbf{x} \leq \mathbf{x}_{ub} \\ & \mathbf{y}_{lb} \leq \mathbf{x} \cdot \mathbf{O} \leq \mathbf{y}_{ub} \\ & \sum_{i \in P_1} E_i \cdot Mw_i = f \cdot E_{tot} \\ & \sum_{i \in P_2} E_i \cdot Mw_i = (1-f) \cdot E_{tot} \\ & \mathbf{N}^* \cdot \mathbf{v} = 0 \end{aligned}$$

P_i are vectors containing the indices of the reactions that belong to each of the pathways, Mw is the vector of molecular weights of the enzyme complexes in Da, as above. The reduced stoichiometry matrix $\mathbf{N}^*$ contains only the rows corresponding to the internal metabolites of both pathways (marked in color in Fig. 7.1 B)
Note that the steady state constraint, the observables constraint as well as the objective function are nonlinear, with the flux vector $\mathbf{v}$ being calculated following the kinetic formulation in Equation 7.1.

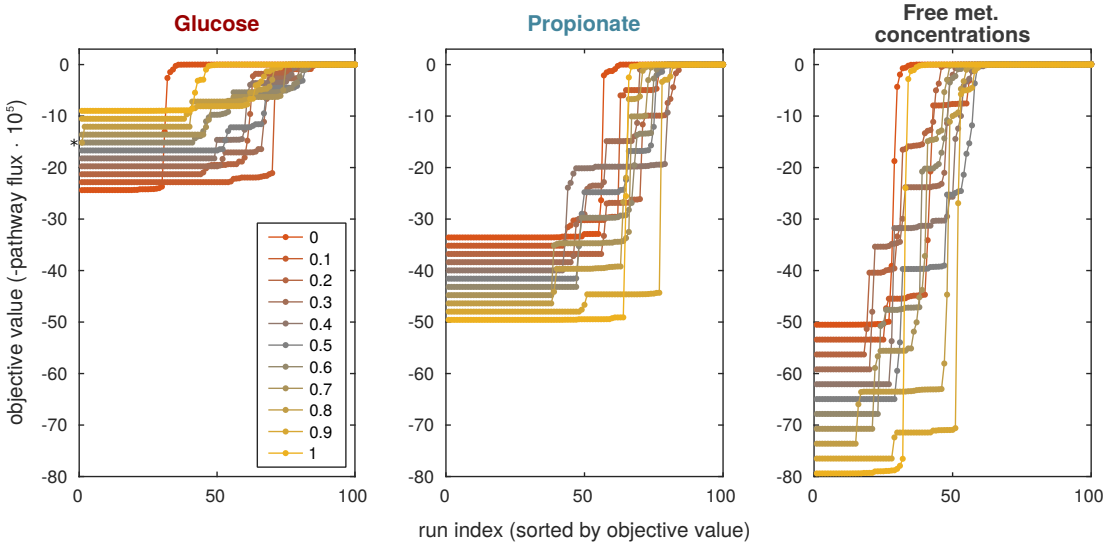


Figure A.28: Convergence profiles of the protein fraction optimization. Colors correspond to f, the percentage of total protein in the MCIT pathway. Please note that for the 0.8 fraction of the glucose, the best fit (marked by an asterisk) was neglected, as it was most likely error prone (deviates from the remaining fits in the global optimum).

A.29 Overview of reactions in Pdtx

Table A.29: Summary of the Pdtx model reactions. Species names are as in Table A.18, $\Delta_r G^0$ and their standard deviations are from eQuilibrator [58], the rate law indices link to Table A.22, *sd* denotes the standard deviation

ID	Reaction name	Reaction	KEGG	EC number	$\Delta_r G^0$ [kJ/mol]	$\Delta_r G^0$ *sd* [kJ/mol]	rate law
pcc	propionyl-CoA carboxylase	ATP + CO2 + PROPIONYLCOA ↔ ADP + PI + SMMALONYLCOA	R01859	6.4.1.3	-5.0	5.3	2
rmcoa	methylmalonyl-CoA epimerase	SMMALONYLCOA ↔ RMMALONYLCOA	R02765	5.1.99.1	0.0	5.8	2
mcm	methylmalonyl-CoA mutase	RMMALONYLCOA ↔ SUCCOA	R00833	5.4.99.2	-7.6	2.1	2
scs	succinyl-CoA synthase	ADP + PI + SUCCOA ↔ ATP + COA + SUCC	R00405	6.2.1.5	1.3	1.4	2
mcs	2-methylcitrate synthase	H2O + OA + PROPIONYLCOA ↔ COA + 2METHYLCITRATE	R00931	2.3.3.5	-42.9	3.4	2
mcd	methylcitrate dehydratase	2METHYLCITRATE ↔ H2O + METHYLACONITATE	R04424	4.2.1.79	-0.9	5.6	2
macn	aconitate hydratase	H2O + METHYLACONITATE ↔ METHYLISOCITRATE	R04425	4.2.1.99	0.9	5.6	2
mcl	isocitrate lyase (methyl)	METHYLISOCITRATE ↔ PYR + SUCC	R00409	4.1.3.30	9.8	3.0	2
RUP_PRO	propionate uptake	→ PROPIONYLCOA	-	-	-	-	5
RBIO_SMMALONYLCOA	Biomass drain for S-Methlymalonyl-CoA	SMMALONYLCOA → BIO_SMMALONYLCOA	-	-	-	-	4
RBIO_PROPIONYLCOA	Biomass drain for Propionyl-CoA	PROPIONYLCOA → BIO_PROPIONYLCOA	-	-	-	-	4

A.30 Overview of enzyme complexes in Pdtx

Table A.30: Enzyme complexes of the Pdtx model, their composition and their subunit concentrations in the two used experimental conditions in fmol/mLOD$_{600}$=1. Abbreviations: Mw - Molecular weight, AA - number of amino acids

ID	RvNumber	subunits/ active site	Mw [Da]	AA [#]	Uniprot ID	gene names	Comments / references	GC2	std	PRO	std
pcc	Rv3285	1	63783	600	P96890	accA3	Alpha-subunit, [62]	1772.4	86.8	1718.5	195.6
	Rv3280	1	59354	548	P9WQH7	accD5	Beta-subunit	573.1	41.8	635.9	7.6
	Rv3281	3	19014	177	P96886	accE5	Eps-subunit	328.6	28.7	381.1	10.1
rmcoa	Rv1322A	2	16626	152	L7N6B1	Rv1322A	Annotation unclear, not in KEGG, [173]	56.9	5.1	66.0	18.0
mcm	Rv1492	1	64744	615	P9WJK7	mutA	[173]	171.4	17.0	164.2	41.1
	Rv1493	1	80604	750	P9WJK5	mutB		181.1	16.7	167.4	29.7
scs	Rv0952	1	31229	303	P9WGC7	sucD	Brenda	2138.2	277.5	1846.5	14.4
	Rv0951	1	40926	387	P9WGC5	sucC		2228.1	215.2	1863.0	50.4
mcs	Rv1131	2	42969	393	I6Y9Q3	prpC	in KEGG as 2.3.3.1	10.3	1.6	553.8	60.1
mcd	Rv1130	1	57783	526	O06582	prpD	Brenda	307.7	128.4	868.9	23.7
macn	Rv1475c	1	102449	943	O53166	acn	in KEGG only as 4.2.1.3, [10]	2279.2	190.2	1907.2	56.8
mcl	Rv0467	4	47087	428	P9WKK7	icl	function carried out by icl1, [67]	343.8	42.3	1216.3	59.4

A.31 Fluxes of Pdtx

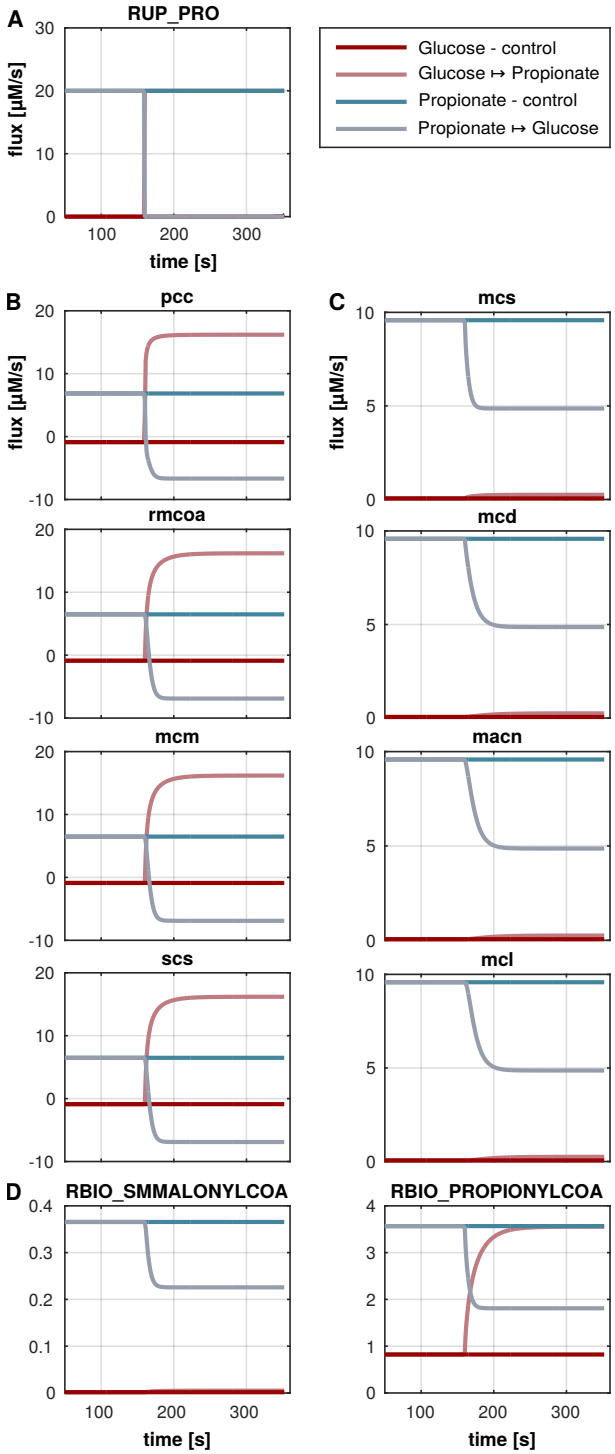


Figure A.31: Fluxes of the Pdtx model corresponding to the simulations shown in Figure 7.5 in μM/s.

A.32 Profile likelihoods of Pdtx model parameters

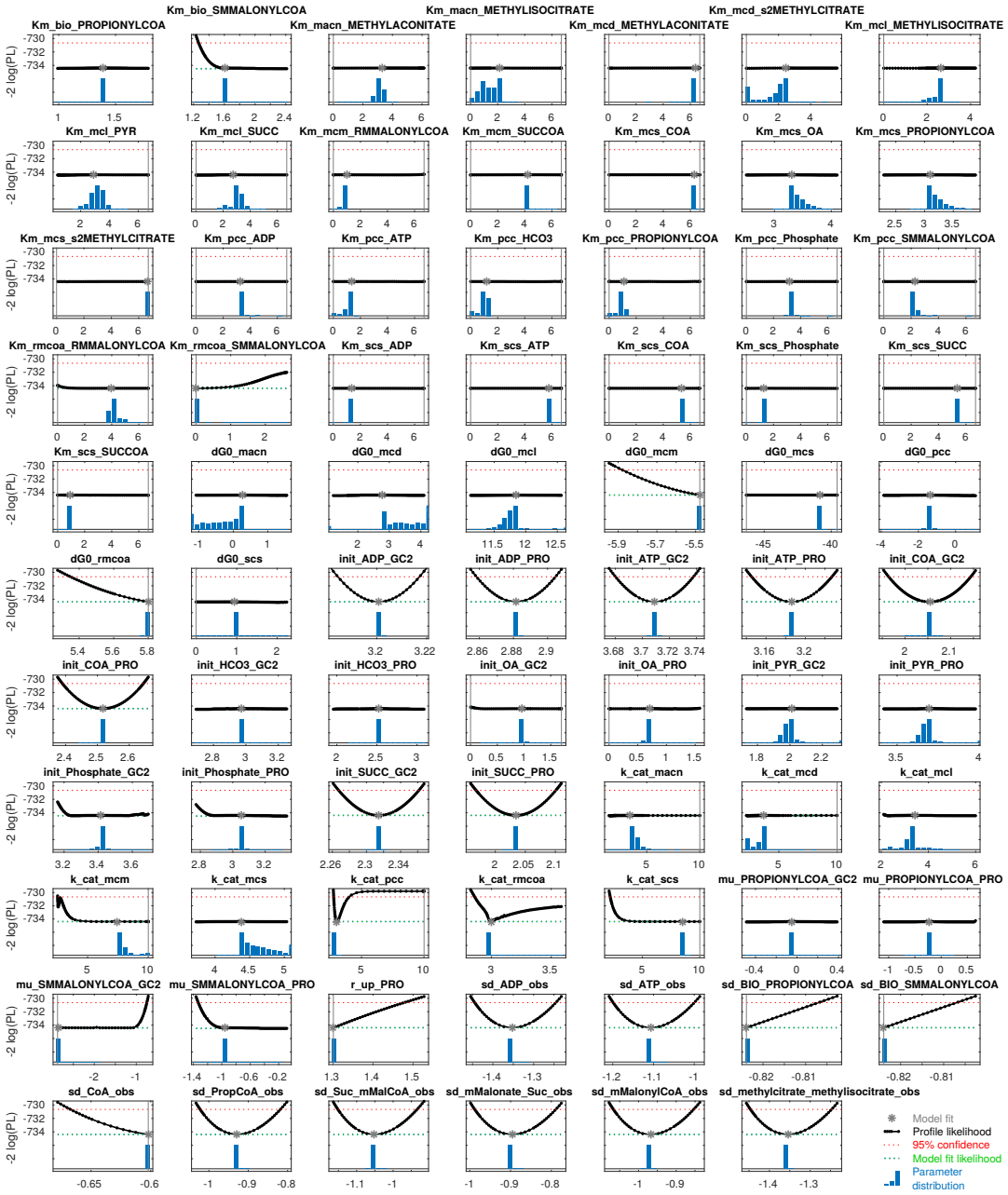


Figure A.32: Summary of the profile likelihoods of all model parameters of the Pdtx model. Each black dot represents one fit with the value of the current parameter fixes to the corresponding value on the x-axis. The y-axis depicts the negative log-likelihood and the red dotted line indicates the 95 % threshold as explained in the Methods. The gray asterisk denoted the found optimum parameter value, the blue histogram includes all values of the the parameter in all profile-likelihood fits (all black dots in the figure). The x-axis is on a logarithmic scale, except for the $\Delta_r G^0$ parameters.

A.33 Parameter values and confidences of Pdtx

Table A.33: Summary of the profile likelihood estimation. For each parametr, the optimum value from the fitting and the lower and upper bounds from the optimization are given. The column islog contains a 1 if the parameter is fitted in log space and a 0 for linear parameters. In the identifiability column, str.nID denotes structural non- identifiability, pra.nID denotes practical non-identifiability and id identifiable parameters. lPl and uPL are resulting lower and upper limits for the confidence intervals. $q_{2.5}$ and $q_{97.5}$ are the 2.5 % and 97.5 % quantiles of the blue distributions shown in Figure A.32.

#	Parameter	fitted value	lb	ub	islog	Identifiability	lPL	uPL	$q_{2.5}$	$q_{97.5}$
1	Km_bio_PROPIONYLCOA	1.39	0	6.7	1	str.nID	$-\infty$	$+\infty$	1.38	1.39
2	Km_bio_SMMALONYLCOA	1.62	0	6.7	1	pra.nID	1.266	$+\infty$	1.61	1.62
3	Km_macn_METHYLACONITATE	3.33	0	6.7	1	str.nID	$-\infty$	$+\infty$	2.61	3.33
4	Km_macn_METHYLISOCITRATE	2.10	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.15	2.10
5	Km_mcd_METHYLACONITATE	6.34	0	6.7	1	str.nID	$-\infty$	$+\infty$	6.34	6.34
6	Km_mcd_s2METHYLCITRATE	2.45	0	6.7	1	str.nID	$-\infty$	$+\infty$	4.08E-05	2.45
7	Km_mcl_METHYLISOCITRATE	2.63	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.30	2.67
8	Km_mcl_PYR	2.92	0	6.7	1	str.nID	$-\infty$	$+\infty$	2.09	3.92
9	Km_mcl_SUCC	2.74	0	6.7	1	str.nID	$-\infty$	$+\infty$	1.91	3.76
10	Km_mcm_RMMALONYLCOA	1.04	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.56	1.04
11	Km_mcm_SUCCOA	4.22	0	6.7	1	str.nID	$-\infty$	$+\infty$	4.22	4.46
12	Km_mcs_COA	6.30	0	6.7	1	str.nID	$-\infty$	$+\infty$	6.30	6.30
13	Km_mcs_OA	3.30	0	6.7	1	str.nID	$-\infty$	$+\infty$	3.28	3.84
14	Km_mcs_PROPIONYLCOA	3.09	0	6.7	1	str.nID	$-\infty$	$+\infty$	3.07	3.65
15	Km_mcs_s2METHYLCITRATE	6.68	0	6.7	1	str.nID	$-\infty$	$+\infty$	6.68	6.68
16	Km_pcc_ADP	3.27	0	6.7	1	str.nID	$-\infty$	$+\infty$	3.27	3.64
17	Km_pcc_ATP	1.39	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.010	1.39
18	Km_pcc_HCO3	1.19	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.008	1.18
19	Km_pcc_PROPIONYLCOA	1.10	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.008	1.10
20	Km_pcc_Phosphate	3.18	0	6.7	1	str.nID	$-\infty$	$+\infty$	3.17	3.54
21	Km_pcc_SMMALONYLCOA	2.25	0	6.7	1	str.nID	$-\infty$	$+\infty$	2.25	2.73
22	Km_rmcoa_RMMALONYLCOA	3.93	0	6.7	1	pra.nID	$-\infty$	$+\infty$	3.93	5.48
23	Km_rmcoa_SMMALONYLCOA	1.08E-05	0	6.7	1	pra.nID	$-\infty$	$+\infty$	2.39E-14	1.08E-05
24	Km_scs_ADP	1.35	0	6.7	1	str.nID	$-\infty$	$+\infty$	1.35	1.35
25	Km_scs_ATP	5.77	0	6.7	1	str.nID	$-\infty$	$+\infty$	5.76	5.76
26	Km_scs_COA	5.37	0	6.7	1	str.nID	$-\infty$	$+\infty$	5.37	5.37
27	Km_scs_Phosphate	1.27	0	6.7	1	str.nID	$-\infty$	$+\infty$	1.27	1.27
28	Km_scs_SUCC	5.36	0	6.7	1	str.nID	$-\infty$	$+\infty$	5.36	5.36

#	Parameter	fitted value	lb	ub	islog	Identifiability	lPL	uPL	$q_{2.5}$	$q_{97.5}$
29	Km_scs_SUCCOA	0.93	0	6.7	1	str.nID	$-\infty$	$+\infty$	0.93	0.93
30	dG0_macn	0.25	-4.7	6.5	0	str.nID	$-\infty$	$+\infty$	-4.58	0.25
31	dG0_mcd	2.77	-6.5	4.7	0	str.nID	$-\infty$	$+\infty$	-0.68	4.70
32	dG0_mcl	11.85	6.8	13	0	str.nID	$-\infty$	$+\infty$	11.47	12.61
33	dG0_mcm	-5.48	-9.7	-5.5	0	pra.nID	-5.89	$+\infty$	-5.48	-5.48
34	dG0_mcs	-40.87	-46	-40	0	str.nID	$-\infty$	$+\infty$	-40.87	-40.70
35	dG0_pcc	-1.38	-15	5.3	0	str.nID	$-\infty$	$+\infty$	-1.87	-0.84
36	dG0_rmcoa	5.80	-5.8	5.8	0	pra.nID	5.34	$+\infty$	5.80	5.80
37	dG0_scs	0.95	-0.0086	2.7	0	str.nID	$-\infty$	$+\infty$	0.64	1.37
38	init_ADP_GC2	3.20	0	4.7	1	id	3.18	3.22	3.20	3.20
39	init_ADP_PRO	2.88	0	4.7	1	id	2.86	2.91	2.88	2.88
40	init_ATP_GC2	3.71	0	4.7	1	id	3.68	3.74	3.71	3.71
41	init_ATP_PRO	3.19	0	4.7	1	id	3.15	3.23	3.19	3.19
42	init_BIO_PROPIONYLCOA	0.00	0	1	1	fixed				
43	init_BIO_SMMALONYLCOA	0.00	0	1	1	fixed				
44	init_COA_GC2	2.05	0	4.7	1	id	1.96	2.15	2.05	2.05
45	init_COA_PRO	2.52	0	4.7	1	id	2.39	2.65	2.52	2.52
46	init_H2O_GC2	6.00	6	6	1	fixed				
47	init_H2O_PRO	6.00	6	6	1	fixed				
48	init_HCO3_GC2	2.97	0	4.7	1	str.nID	$-\infty$	$+\infty$	2.88	3.11
49	init_HCO3_PRO	2.53	0	4.7	1	str.nID	$-\infty$	$+\infty$	2.44	2.65
50	init_METHYLACONITATE_GC2	0.56	0	4.7	1	fixed				
51	init_METHYLACONITATE_PRO	2.07	0	4.7	1	fixed				
52	init_METHYLISOCITRATE_GC2	0.46	0	4.7	1	fixed				
53	init_METHYLISOCITRATE_PRO	1.93	0	4.7	1	fixed				
54	init_OA_GC2	0.96	0	1.7	1	str.nID	$-\infty$	$+\infty$	0.92	1.05
55	init_OA_PRO	0.72	0	1.7	1	str.nID	$-\infty$	$+\infty$	0.56	0.73
56	init_PROPIONYLCOA_GC2	1.36	0	4.7	1	fixed				
57	init_PROPIONYLCOA_PRO	2.16	0	4.7	1	fixed				
58	init_PYR_GC2	2.01	0	4.7	1	str.nID	$-\infty$	$+\infty$	1.91	2.46
59	init_PYR_PRO	3.70	0	4.7	1	str.nID	$-\infty$	$+\infty$	3.60	4.13
60	init_Phosphate_GC2	3.41	0	4.7	1	pra.nID	$-\infty$	$+\infty$	3.36	3.49
61	init_Phosphate_PRO	3.05	0	4.7	1	pra.nID	$-\infty$	$+\infty$	2.99	3.14
62	init_RMMALONYLCOA_GC2	0.67	0	4.7	1	fixed				
63	init_RMMALONYLCOA_PRO	1.01	0	4.7	1	fixed				

#	Parameter	fitted value	lb	ub	islog	Identifiability	lPL	uPL	$q_{2.5}$	$q_{97.5}$
64	init_SMMALONYLCOA_GC2	1.67	0	4.7	1	fixed				
65	init_SMMALONYLCOA_PRO	2.14	0	4.7	1	fixed				
66	init_SUCCOA_GC2	1.63	0	4.7	1	fixed				
67	init_SUCCOA_PRO	1.97	0	4.7	1	fixed				
68	init_SUCC_GC2	2.31	0	4.7	1	id	2.27	2.36	2.31	2.31
69	init_SUCC_PRO	2.03	0	4.7	1	id	1.97	2.10	2.03	2.03
70	init_s2METHYLCITRATE_GC2	0.97	0	4.7	1	fixed				
71	init_s2METHYLCITRATE_PRO	2.44	0	4.7	1	fixed				
72	k_cat_macn	3.39	-12	10	1	str.nID	$-\infty$	$+\infty$	3.39	4.89
73	k_cat_mcd	3.71	-12	10	1	str.nID	$-\infty$	$+\infty$	2.15	3.85
74	k_cat_mcl	3.43	-12	10	1	str.nID	$-\infty$	$+\infty$	2.31	4.42
75	k_cat_mcm	7.42	-12	10	1	pra.nID	2.53	$+\infty$	7.42	9.95
76	k_cat_mcs	4.38	-12	10	1	str.nID	$-\infty$	$+\infty$	4.34	5.54
77	k_cat_pcc	3.04	-12	10	1	id	2.77	4.08	3.02	3.06
78	k_cat_rmcoa	3.00	-12	10	1	pra.nID	2.86	$+\infty$	2.99	3.03
79	k_cat_scs	8.46	-12	10	1	pra.nID	2.18	$+\infty$	8.45	8.51
80	mu_PROPIONYLCOA_GC2	-0.05	-2.9	0.8	1	str.nID	$-\infty$	$+\infty$	-0.07	-0.05
81	mu_PROPIONYLCOA_PRO	-0.22	-2.9	0.8	1	str.nID	$-\infty$	$+\infty$	-0.23	-0.21
82	mu_SMMALONYLCOA_GC2	-2.86	-2.9	0.8	1	pra.nID	$-\infty$	-0.75	-2.86	-2.84
83	mu_SMMALONYLCOA_PRO	-0.96	-2.9	0.8	1	pra.nID	-1.33	$+\infty$	-0.96	-0.95
84	r_up_PRO	1.30	1.3	2.7	1	pra.nID	$-\infty$	1.48	1.30	1.30
85	sd_ADP_obs	-1.35	-5	-0.6	1	id	-1.45	-1.24	-1.35	-1.35
86	sd_ATP_obs	-1.11	-5	-0.6	1	id	-1.20	-1.00	-1.11	-1.11
87	sd_BIO_PROPIONYLCOA	-0.82	-0.82	-0.4	1	pra.nID	$-\infty$	-0.81	-0.82	-0.82
88	sd_BIO_SMMALONYLCOA	-0.82	-0.82	-0.4	1	pra.nID	$-\infty$	-0.81	-0.82	-0.82
89	sd_CoA_obs	-0.60	-5	-0.6	1	pra.nID	-0.67	$+\infty$	-0.60	-0.60
90	sd_PropCoA_obs	-0.93	-5	-0.6	1	id	-1.02	-0.82	-0.93	-0.92
91	sd_Suc_mMalCoA_obs	-1.05	-5	-0.6	1	id	-1.15	-0.93	-1.05	-1.05
92	sd_mMalonate_Suc_obs	-0.90	-5	-0.6	1	id	-0.99	-0.79	-0.90	-0.90
93	sd_mMalonylCoA_obs	-0.97	-5	-0.6	1	id	-1.07	-0.85	-0.97	-0.96
94	sd_methylcitrate_methylisocitrate_obs	-1.35	-5	-0.6	1	id	-1.45	-1.25	-1.36	-1.35

A.34 Response coefficients of Pdtx

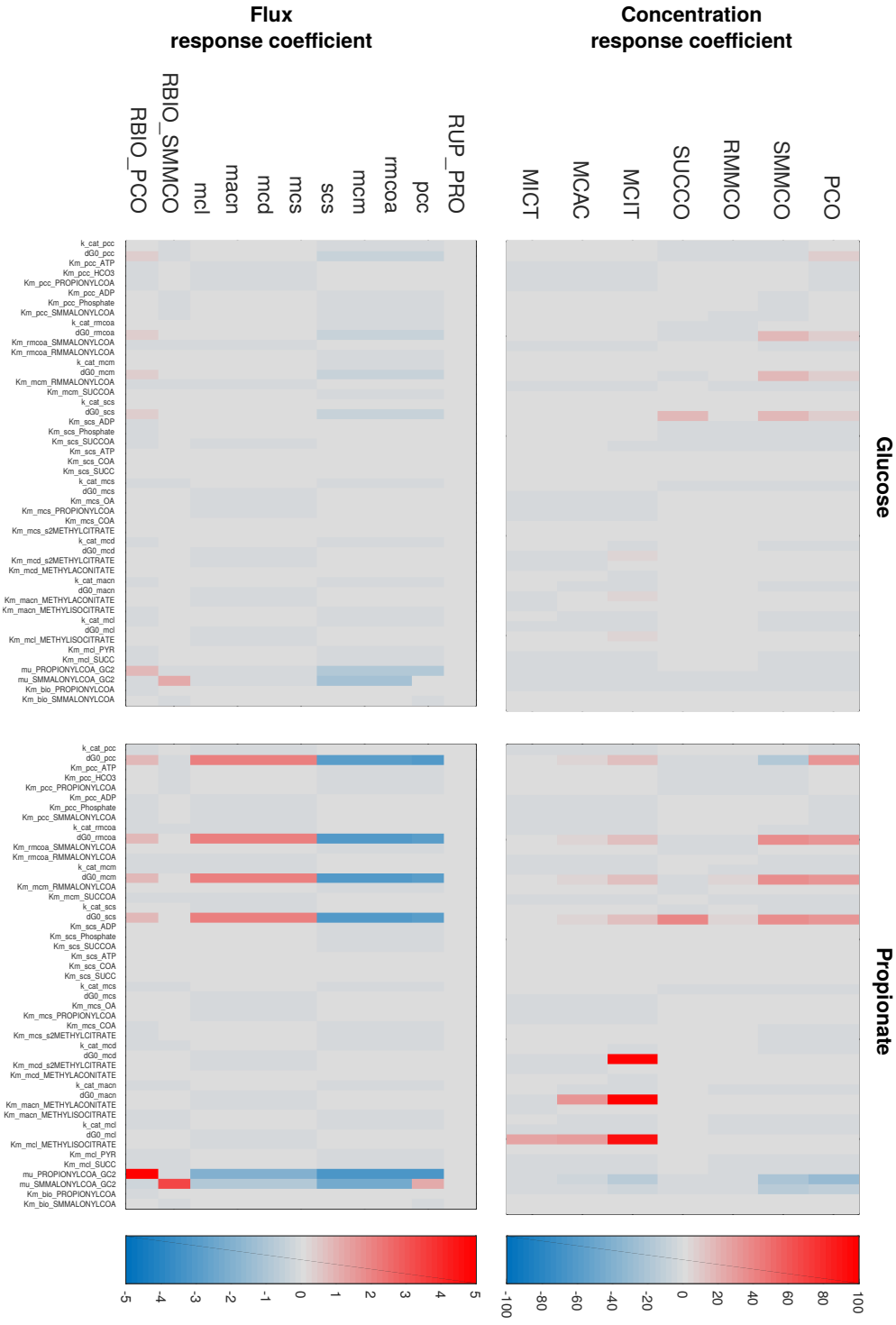


Figure A.34: Response coefficients of the Pdtx model, corresponding to the simulations in Figure 7.5.

A.35 Example simulation of the tCCMtb model

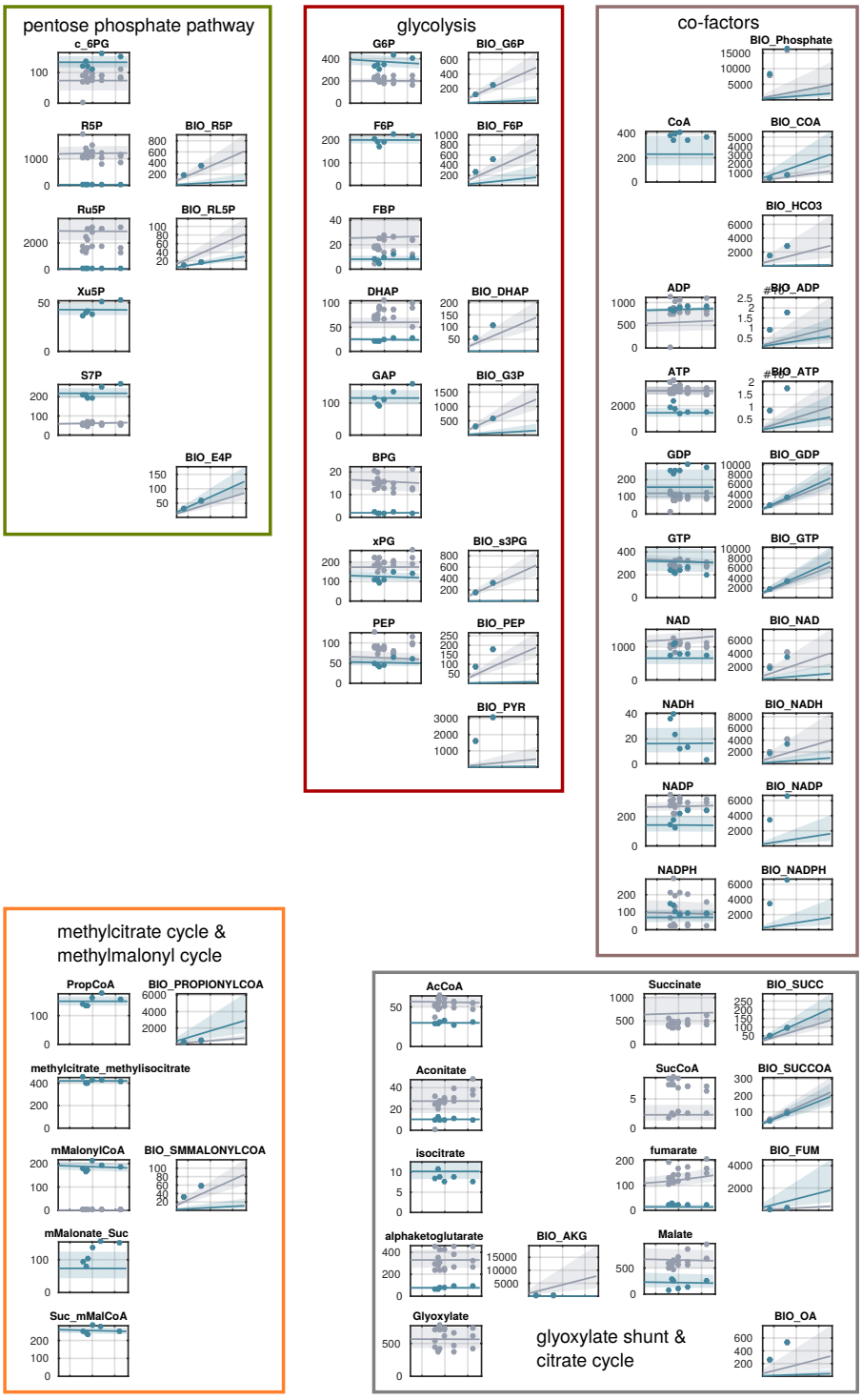


Figure A.35: Simulations and data of two individual model simulations (lines and shaded error estimates) for propionate (blue) and glutamate (gray) control data (dots).

The Road goes ever on and on,
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can.

Bilbo Baggins